

[illegible]

XX The present sequence encodes a human inorganic pyrophosphatase CC designated HPYP. Nucleic acids encoding HPYP were first identified CC in Incyte clone 768320 from the lung tissue cDNA library LUNGNOT04 CC CC The recombinant pyrophosphatase, may be useful in promoting cell CC and tissue regeneration..

Sequence 1275 BP; 394 A; 235 C; 294 G; 351 T; 1 other;

Query Match	99.9%	Score 1274;	DB 20;	Length 1275;
Best Local Similarity	100.0%	Pred. No: 0;		
Matches 1275: Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0;

QY	1	CAAGAGTTGGGGCTCTCTCCTTTCAGTCGGGCGCGCGTGGGGCTGGTGGCTCTGTG	60
Db	1	caaaagctcttggagctctctccctctgtcaagctcgcgccgctgctggctgtctctgtg	60
QY	61	GCAAGCGGCGGCGAGGACCTCGGCACATGAGCGGCTTAGCAGCGAGAGCGCGCG	120
Db	61	gcagcgcgcgcgaggaagctccgcacatactgagcggtcttcagaccgagagcgcgcg	120
QY	121	GCCCTTCCCTCGAGTACCGAGCTTCCCTCAAAAATGAGAAGGACAAATATATCTCC	180
Db	121	gcccttcccttcgagtaaccgagcttccccaataaagagaagagcaataatactcc	180
QY	181	ATTTCATGATATTCCAATTTATGTCAGATTAAGATGTGTTTCACATGGTAGTTGAAGACC	240
Db	181	atttcagatattccaatattatgacgaataagatgtgttccaatgttagttagaagttacc	240
QY	241	ACCGTGCTTAATGCAAAAATGAGATGTGCTACAAAGACCCCTTTAAACCTTTAAACA	300
Db	241	acgcgtgtcttaatgcaaaaaatgagatgtgtctacaagaagaccctttaacccttaaca	300
QY	301	AGATGTAAAAAAGGAAACCTCGCTATGTGGGAATTTTCCCGATTAAGATATAT	360
Db	301	agatgtgaaaaaaggaaaacactcgctatgttgcgaattgttcccgataaagatatat	360
QY	361	CTGGAACATATGTTGCCATCTCCCTAGACATTGGGAAGACCAGGCGCACATATTAACATAC	420
Db	361	ctggaacatattgtgtccatctccctcgaacttgggaagaccagggacaatgataaactac	420
QY	421	TGGCTGTTGTGTGATACATACCAACCAATGTATGTGTGTAATTGGAAATTGGAAGCAAGTATGTGC	480
Db	421	tggctgttgtgtgtgacatatacccaattgtatgtgtgtgaatttggaagcaagtatgtgc	480
QY	481	AAGAGTGAATAATTGGCGGTGAAGTTTACGCAATTTGGCTATGATTTACGAAGGGGA	540
Db	481	aagagtgaaataatttggcgttgaaagttttagcgaattttggctatgattttacgaaggga	540
QY	541	AACCGATGCAAAAGTCATTTGCCATTATGTGGATGCGATGACACCAATTTATATGA	600
Db	541	aaccgactgaaaagctacttgcacataatgtgatatgtccctgagcgacattataatga	600
QY	601	TATCATGATGTCAAAAGCGCTGAACCTGGCTACTTAGAAGTACTGTGTGCTGTTAG	660
Db	601	tatcataatgatacaaaagcgctgaaccctgctaacttgaagctactgtgactggtttag	660
QY	661	AAGTATTAAGGTTCTCTATGTAAGAAACAGAAATGAGTTTGCTTTATCTACAATTAA	720
Db	661	aaggtataaggtctctcgtatgtaaaaaccgaaaatgagtttggttttaattgcagaatttaa	720
QY	721	AGATTAAGCACTTGGCCATTGATATTTATTAANAAGCACTCATGACCATTGGAAGCAATTAGT	780
Db	721	agataaagcacttgcctatgtatataataaaacacatactgacacattggaagaacattagt	780
QY	781	GACTTAACAAAACGATATGTAAGAAAGAAATCACTTGCAATTAATACACTTTGCTGAGAGCC	840
Db	781	gacttaagaaaacgaattgtaaaaaggaatcagttgcatagaatacaacttgcctggaagccc	840
QY	841	CTTCAAGTGTGATCTTGATGCTCGCACAGGCATTGTGTGATGTCTTTACACACACCTGTGA	900
Db	841	cttcaaatgtgatacccttaactctgcaaaagcattgtgatagtcttaccacacaccttata	900

OY	901	ATTCGCTGCACAGTACCAACAGACGGATTAAGTGGTTCATCAACGAAAAACTAATG	960
Db	901	atctgcctgcacag taccacacagacgcygga taagtggtccatccaccgaaaaactaatg	960
OY	961	AGATTTTCTTGGGAATTAACACTGATATTGCTACATCGTGTTCATCTGGATGTATTAGAG	1020
Db	961	agattcttcygaatacaaacgtatcatctgcatcatcggtcatctcygactgtattagaag	1020
OY	1021	TAAAGTAGTAGCTTTTCAAGCGTTTAAATTTTATGACATCACTCTTACCAAGTAAATTC	1080
Db	1021	taaaagtagtagctctttccaagctttaaattcttgtagaacctcatcaactaaagtaaatc	1080
OY	1081	TGCTGTGACTAATCAATATATCTCAGAAATGTTATTCATCTTAAAGCATTTTTCATATTC	1140
Db	1081	tgctgtgactaaatccaatatactacagaagtatcatcatctaaagcatctttcatatctca	1140
OY	1141	ACTAAGATTAACCTTTAGCACATGCTTAATATCTAAAGCAGTGTCTATTGGGAAGTCACTT	1200
Db	1141	actaagaataactctttagcacatgctttaaatatacaaacgactgtgcaattttagaagcaactt	1200
OY	1201	GTTGAATAGATATGCAAGGGAGACACATATTTGATGTATTTTACCATPTATTAGAGAAAT	1260
Db	1201	gttgaatagaatgtgcaagggagacacatatgtgatgtatattgttaccacatatgttgaagaat	1260
OY	1261	AAAATATATTTGCTG	1275
Db	1261	aaaattatttgcctg	1275

RESULT	3
AAS08596	
ID	AAS08596 standard; cDNA; 1277 BP.

AC AAS08596

DT 24-OCT-2001 (first entry)

DE DNA encoding human pyrophosphatase (HPYP).

KW Pyrophosphatase; human; HPYP; antiinflammatory; cytostatic;

KW antirheumatic; antiarthritic; inflammatory disease; cancer; nerve;
antihypertensive; anticoagulant; antidiabetic; osteoporosis

KW systemic lupus erythematosus: myasthenia gravis: diabetes mellitus: kw transplacental; pancreatic; bone marrow; anaemia; ascites;

kw scleroderma: ss: osteoporosis; giointerstitialis; rheumatoid arthritis; osteoarthritis

XX
XX
Homo sapiens

XX	XX	Key
XX	XX	Key

Location/Qualifiers

CDS
ET

```

91.:900
/*tag= a

```

XXI

pyriophosphate-
human

	XX
IV YTCOTATCZCO	N7

XX	0
XX	1
XX	2
XX	3
XX	4
XX	5
XX	6
XX	7
XX	8
XX	9

XX

PR 13-AUG-1998;

98US-0134593.

PA (HAWK/) HAWK:

INS P R.

XX

XX

DR P-PSDB; AAU03

5122.

PT New human py

rophosphatase polypeptides and polynucleotides, useful for

Claim 5; Fig 1; 32pp; English

XX Sequence 1277 BP; 394 A; 235 C; 294 G; 351 T; 3 other;
SO

1 CAAGAGTTNGGGCTCTCTCCTTGTCAAGTCGGCGCCGGCTGCGGGCTGGTGCTCTC

Db 3 caagaggttngggctctctactgttcagtcgagccgctgacggctcgttgcgtctct

100

XX
AC
AAX00685;

XX Human; secreted protein; fusion protein; gene
KW

XX
05 Homo sapiens.

XX

PF 19-MAR-1998; 98WO-US05311.
 XX 30-MAY-1997; 97US-0050937.
 PR 21-MAR-1997; 97US-0041276.
 PR 21-MAR-1997; 97US-0041277.
 PR 21-MAR-1997; 97US-0041281.
 PR 21-MAR-1997; 97US-0042344.
 PR 30-MAY-1997; 97US-0048069.
 PR 30-MAY-1997; 97US-0048094.
 PR 30-MAY-1997; 97US-0048095.
 PR 30-MAY-1997; 97US-0048096.
 PR 30-MAY-1997; 97US-0048099.
 PR 30-MAY-1997; 97US-0048131.
 PR 30-MAY-1997; 97US-0048135.
 PR 30-MAY-1997; 97US-0048154.
 PR 30-MAY-1997; 97US-0048160.
 PR 30-MAY-1997; 97US-0048186.
 PR 30-MAY-1997; 97US-0048187.
 PR 30-MAY-1997; 97US-0048188.
 PR 30-MAY-1997; 97US-0048350.
 PR 30-MAY-1997; 97US-0048351.
 PR 30-MAY-1997; 97US-0048352.
 PR 30-MAY-1997; 97US-0048355.
 PR 05-AUG-1997; 97US-0054804.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Brewer LA, Duan R, Ebner R, Fertle AM, Florence KA;
 PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
 PI Rosen CM, Ruben SM, Shi Y, Young P;
 XX WPI: 1999-070066/06.
 DR P-PSDB: AAM67881.
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 1; Page 233-234; 385pp; English.
 XX
 CC This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number, and the clone it is derived
 CC from, are detailed in the descriptor line. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human immunoglobulin
 CC Fc portion (e.g. AAX00602) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 87 novel genes and their fragments (nucleic
 CC acid sequences: AAX00611-X00724; amino acid sequences AAM67807-W68004)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 87
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX00611 for described uses).
 CC
 XX Sequence 1351 BP; 414 A; 252 C; 319 G; 362 T; 4 other;
 SQ

Db 134 ttcccttgagtagaccgagcttccctcaaaaagagaagacatatatactccatt 193
 Oy 185 CATGATATTCGAATTTATGACGATGAAGGATGTTTCAAGTAGTGAATGACACGC 244
 Db 194 catgatatcccaattcagcagataaagatgcttccacatgagtagtgaagtagcaccgc 253
 Oy 245 TGGCTTAATGCAAAAATAGAGATGCTCTCAAAAGACCCCTTAACCCCTATTAAACAAGAT 304
 Db 254 tgcictaatgcaaaaatgagagattgctacaaagaccctttaaaccctttaaacaagat 313
 Oy 305 GTGAAAAAAGCAAAACCTTCGTATGTTGCGAATTTGTCCTGATTAAGATATATCTGCG 364
 Db 314 gtgaaaaaagaaaacattcgtatgctggaattgttcccgatataagagatatattcgg 373
 Oy 365 AACATGCTGCATCCCTCAGACCTTGGAAGACCCAGGACCAATGATTAACATCTGCG 424
 Db 374 aacctatgctccatccctccagacttggaaagaccagagccaatgataaacatctggc 433
 Oy 425 TGTGTGTGACAAATGACCAATTGATGTGTGAATTTGGAAGCAAGATATGTCAGA 484
 Db 434 tgtgtgtgacaaatgaccctaatgtatgtgtgaatttggaaggaagtatgtgcaaga 493
 Oy 485 GGTGAATTAATTTGGCTGAAGCTTACGCAATTTGCTATGATTTGACGAAGGGAAC 544
 Db 494 ggtgaataattggtgtgaagttctaggaatgcatgtgatagtatgacgaagggaagacc 553
 Oy 545 GACGTGAAGATCATTCGCTTAATGATGATGATGATGATGATGATGATGATGATGATG 604
 Db 554 gactggaagatcatgcatctaaatgtagatgacccatgacgaatataaataatc 613
 Oy 605 AATGATGTCAAAACGCTGAACCTGCTACTAGTGAAGCTAGTGTGATGATGATGATG 664
 Db 614 aatgattgcaaaagcgtgaaacccgtgacttagaagtagtactgtgtttagaag 673
 Oy 665 TATAAGTCTCTGATGAAACCAAGAAATGAGTTGCTTAAATGCAATTAATTAAGAT 724
 Db 674 tataagttctctgtagtgaacccaagaaatgattgtgtttaaagcagaatttaaag 733
 Oy 725 AAGGACTTGGCATATGATATATTAATTAAGCACTGATGACCAATTAATTAATGATG 784
 Db 734 aaggacttggcatatgatatatttaaagcaactcagcaacttgaagcatttagtact 793
 Oy 785 AAGAAAGCAATGGAAGGAATGATGATGATGATGATGATGATGATGATGATGATG 844
 Db 794 aagaaagcaatggaaggaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 853
 Oy 845 AAGTGTATCTGATGCTGCAAGACCAATTTGATGCTTTACCAACCCCTGTGAATCT 904
 Db 854 aagttgtatctgtagtgcagaaagccatttgtatgctttaaaccacccctgtgaact 913
 Oy 905 GCCTGCACAGTACCAACAGAGGTGATGATGATGATGATGATGATGATGATGATGATG 964
 Db 914 gccctgcacagtaaccaagagcgtgataagtggttccatccacaaaacaaagagat 973
 Oy 965 TTCTCTGAATATCAAGCTGATATTTGCTACATCGTCTTCAATGATGATGATGATGAT 1024
 Db 974 ttctctgaaatacaagcgcgataattgctacatcggttcatcgtgtatagaaagtaa 1033
 Oy 1025 AGTACTACTTTTCAAGCTTAAATTTGTAGACTCATCTTAAGTAATATTTCTGCT 1084
 Db 1034 agtactacttttcaagcgtttaaatttgtagaactcttaactaaagtaaatctgct 1093
 Oy 1085 GTGACTAATCCAAATATCTACGAATGTTATTCATGATTAAGCAATTTTTCATATCTCACTA 1144
 Db 1094 gtgactaatccaaatattactcgaatgtatcatcaatlaaagaatttcatatctcaacta 1153
 Oy 1145 AGATAACTTTTATGACATGCTTAATTAATCAAAAGAGTTGCTATTGGAAGTCACTTGTGA 1204
 Db 1154 agataacttttatagcacatgctttaaatacaaaagcgtgtcatcttggaagcacttgga 1213
 Oy 1205 ATGATGTGCAAGGGGACACATATTTGATGATGATGATGATGATGATGATGATGATG 1264
 Db 1214 atgattgtgcaaggaggacacatattgattatattgttaccatattgttgaagaaataaaa 1273

OY	1265	TATATTGCTG	1275
Db	1274	ttaattgcctg	1284
RESULT	5		
ID	AAK00716	standard; DNA; 1350 BP.	
XX	AAK00716;		
AC	25-MAR-1999	(first entry)	
XX			
DE	Human secreted protein gene 75 clone H0SE145.		
XX			
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;		
KM	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia		
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;		
KM	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;		
KW	inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;		
KM	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus		
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyrid; digestion;		
KM	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.		
XX	Homo sapiens.		
OS			
XX	WO9842738-A1;		
PN	01-OCT-1998.		
XX			
PD	19-MAR-1998;	98MO-US05311.	
PF			
XX	30-MAY-1997;	97US-0050937.	
PR	21-MAR-1997;	97US-0041276.	
PR	21-MAR-1997;	97US-0041277.	
PR	21-MAR-1997;	97US-0041281.	
PR	21-MAR-1997;	97US-0042344.	
PR	30-MAY-1997;	97US-0048069.	
PR	30-MAY-1997;	97US-0048094.	
PR	30-MAY-1997;	97US-0048095.	
PR	30-MAY-1997;	97US-0048096.	
PR	30-MAY-1997;	97US-0048099.	
PR	30-MAY-1997;	97US-0048131.	
PR	30-MAY-1997;	97US-0048135.	
PR	30-MAY-1997;	97US-0048154.	
PR	30-MAY-1997;	97US-0048160.	
PR	30-MAY-1997;	97US-0048186.	
PR	30-MAY-1997;	97US-0048187.	
PR	30-MAY-1997;	97US-0048188.	
PR	30-MAY-1997;	97US-0048350.	
PR	30-MAY-1997;	97US-0048351.	
PR	30-MAY-1997;	97US-0048352.	
PR	30-MAY-1997;	97US-0048355.	
PR	05-AUG-1997;	97US-0054804.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX	Brewer LA; Duan R; Ebner R; Ferrie AM; Florence KA;		
PI	Greene JM; Hu JS; Laflaur DM; Moore PA; Ni J; Olsen HS;		
PI	Rosen CA; Ruben SM; Shi Y; Young P.		
XX			
DR	WPI; 1999-070066/06.		
XX	P-PSDB; AAM67912.		
PT	New isolated human genes and the secreted polypeptides they encode -		
PT	useful for diagnosis and treatment of e.g. cancers, neurological		
PT	disorders, immune diseases, inflammation or blood disorders		
PS	Claim 1; Page 264-265; 385pp; English.		
CC	This sequence represents a nucleic acid molecule which encodes a		
CC	secreted human protein. The gene number, and the clone it is derived		

Db 721 aagagcttgcattgatatataaagacatcagcattggaagcattagtgact 790
 Oy 785 AAGAAACGAATGCAAGAAAGATGCTGATGATACACTTGTGAGAGCCCTTC 844
 Db 791 aagaaacgaatggaagaaagatcattgattgttcaacttctgtggccctcaaa 850
 Oy 845 AAGTG 849
 Db 851 gtgtg 855

RESULT 8
 AAC10628
 ID AAC10628 standard; cDNA: 660 BP.
 AC AAC10628;
 XX 06-OCT-2000 (first entry)
 XX Human secreted protein 5' EST, SEQ ID NO: 14703.
 XX Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX Homo sapiens.
 OS EP1033401-A2.
 XX 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-0200610.
 XX 26-FEB-1999; 99US-0122487.
 XX (GIST) GENSET.
 PA Dumas Milne Edwards J, Duclert A, Giordano J;
 DR WPI: 2000-500381/45.
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PS Claim 1: SEQ ID 14703; 71bp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 660 BP: 218 A; 112 C; 133 G; 197 T; 0 other;

Query Match 51.7%; Score 659; DB 21; Length 660;
 Best Local Similarity 100.0%; Pred. NO. 4.5e-177;
 Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 617 CGGCTGAACCTGCTACTAGAACCTACTGTGACTGCTTTAGAGATTAAGCTTCT 676
 Db 1 cggctgaacacctgtactagaagctactgtgactgtgtttagaagtagtcttct 60
 Oy 677 GATGGAACCAAGAAATGAGTTTGGTTTATGCAAGATTAAAGATTAAGACTTTGGC 736

Db 61 gatgaaacacgaaaatgagtttcgttcaatgcaagattaaagacatttcgc 120
 Oy 737 ATTGATATTATTAAGAGACATCATGACATTCGTAAGATTAAGAAAGAT 796
 Db 121 attgattattataaagacatcattgacattggaagatttagtactaaagaaacgaat 180
 Oy 797 GGAAGAAGATCATGTTGCTATGATGAATCACTTGTCTGAGAGCCCTTAAGTGTATCT 856
 Db 181 ggaagaagaaatcagtttgatgaatacaacttgcctgagagcccttaagtgtatcct 240
 Oy 857 GATGCTGCCAGACCATGTTGATGCTTTACCCACCACCCCTGTGAATTCCTGCACAGTA 916
 Db 241 gatgtccagagacattgtgatgtcctttaccacacccttgatcctgcctgcacagta 300
 Oy 917 CCAACAGACGTGATTAAGTGTTCATCATACCAAGAAATTAATGATTTCTCTGATA 976
 Db 301 ccaacagacgttgataagtggttccatcacacgaagaaactaagattctcttggaata 360
 Oy 977 CAAGCTGATATTGCTACATCGTTCATCTGATGATGATTAAGTAAGTAGTACTTT 1036
 Db 361 caagctgatatgtctacatcgttgcattcctgagatgataagaaagtagtactt 420
 Oy 1037 TCAAGCTTTAAATTTGATGAACTCATCTAACTAAAGTAAATTCGTCTGATTAATCCA 1096
 Db 421 tcaagctttaaatlttgatgaactcatcactaaactaaatltcgtcgtgactaatcca 480
 Oy 1097 ATATCTCAGATGTTATCCATCTTAACATTTTTCATATCATCACTAAGTAAGTAATTTTA 1156
 Db 481 atatactagaatgttattccatcctaagacatttccatcctaagataaacttcta 540
 Oy 1157 GCACATGCTTAAATTCACAGAGAGTTCATTTGGAAGTCACTTGTGAATGATGCA 1216
 Db 541 gcaatgtcttaataataatacaagcagttgcatlttggaagctactgtgatatgagtcgaa 600
 Oy 1217 GGGAGACACATATTGATGATTAATGTTACCATATGTTAGGAATTAATTTTGTCTG 1275
 Db 601 gggagacacatttggatgatatattacacatatgttagaataaattatttgcgtg 659

RESULT 9
 AAA78135/c
 ID AAA78135 standard; cDNA: 578 BP.
 AC AAA78135;
 XX 14-NOV-2000 (first entry)
 XX cDNA encoding human colon tumour polypeptide, SEQ ID NO:422.
 XX Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
 KW immunotherapy; diagnosis; progression; ss.
 OS Homo sapiens.
 PN WO200037643-A2.
 XX 29-JUN-2000.
 XX 23-DEC-1999; 99WO-US30909.
 XX 23-DEC-1998; 98US-0221298.
 XX 02-JUL-1999; 99US-0347496.
 XX 22-SEP-1999; 99US-0401064.
 XX 19-NOV-1999; 99US-0444242.
 XX 02-DEC-1999; 99US-0454150.
 XX (CORI-) CORIXA CORP.
 XX Xu J, Lodés MJ, Secretist H, Benson DR, Meagher MJ, Stolk J;
 PI Wang T, Yuguu J;
 XX WPI: 2000-442671/38.

XX New colon tumor polypeptides used to inhibit the development of cancer,
 PT especially colon cancer, and for diagnosing and monitoring the
 PT progression of the cancer -

PS Claim 1: Page 212; 229pp; English.

XX Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
 CC portions of proteins which are associated with human colon tumors.
 CC The invention also specifically discloses 8 human colon tumor proteins
 CC (AA11887-81104). The nucleic acids, the polypeptides they encode, and
 CC antigen presenting cells (APCs), preferably dendritic cells, expressing
 CC such polypeptides may be used in vaccines that target tumor cells,
 CC especially colon tumor cells, thereby inhibiting the development of
 CC cancer. T-cells specific for the polypeptide expressed by the APC are
 CC used to remove tumor cells from biological samples, especially blood or
 CC fractions thereof. The sample or the isolated T-cells specific for the
 CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
 CC CD8+ T-cells from a patient may be incubated with a polypeptide or
 CC nucleic acid of the invention, or an APC expressing such a polypeptide,
 CC to cause the proliferation of specific T-cells. The T-cells can be
 CC cloned and then administered back to the patient to inhibit cancer
 CC development. Nucleic acids encoding the polypeptides and antibodies
 CC against the polypeptides may be used to determine the expression level
 CC of a tumor protein of the invention, and therefore to determine whether
 CC cancer cells are present. Such diagnostic methods may also be used to
 CC monitor the progression of a cancer by repeating the processes at time
 CC intervals, and comparing the current result to previous results. The
 CC present sequence represents a cDNA encoding a human colon tumor
 CC polypeptide.

XX Sequence 578 BP; 155 A; 139 C; 99 G; 185 T; 0 other;

Query Match 44.0%; Score 560.6; DB 21; Length 578;
 Best Local Similarity 99.1%; Pred. No. 3.6e-149; Indels 1; Gaps 1;
 Matches 574; Conservative 0; Mismatches 4;

XX 337 TTTGTTCCGTTAAAGATATATCTGAACTATGTCATCCCTCAGACTTGGGAGA 396
 DB 578 TTTGTTCCGTTAAAGATATATCTGAACTATGTCATCCCTCAGACTTGGGAGA 519
 XX 397 CCCAGGCAATGATTAACATCTGCTGCTGGTGGACATGACCAATGATGTGTG 456
 DB 518 CCCAGGCAATGATTAACATCTGCTGCTGGTGGACATGACCAATGATGTGTG 459
 XX 457 TGAATTTGAGAGAGATGTGCAAGAGTGAATTAATTTGGCGTGAAGTTCTAGCAT 516
 DB 458 TGAATTTGAGAGAGATGTGCAAGAGTGAATTAATTTGGCGTGAAGTTCTAGCAT 400
 XX 517 ATTGGCTATGATGCAAGAGGGAACCGATGGAAGCATTCGATATATGTGATGA 576
 DB 399 ATTGGCTATGATGCAAGAGGGAACCGATGGAAGCATTCGATATATGTGATGA 340
 XX 577 TCCGATGACGACCAATTAATATGATATCAATGATGCAAAAGCGCTGACTT 636
 DB 339 TCCGATGACGACCAATTAATATGATATCAATGATGCAAAAGCGCTGACTT 280
 XX 637 AGAAGTACTGTGACTGTTTGAAGAGTATAGTTCCTGATGGAAGCAAGAAATGA 696
 DB 279 AGAAGTACTGTGACTGTTTGAAGAGTATAGTTCCTGATGGAAGCAAGAAATGA 220
 XX 697 GTTTCGCTTAATGACAGATTAAGATTAAGACTTGGCATGATATTTAAAGCAC 756
 DB 219 GTTTCGCTTAATGACAGATTAAGATTAAGACTTGGCATGATATTTAAAGCAC 160
 XX 757 TCATGACCATTTGAAGACATTAAGTACTAGCAAGAAACCAATGGAAGATTCAGTGCAT 816
 DB 159 TCATGACCATTTGAAGACATTAAGTACTAGCAAGAAACCAATGGAAGATTCAGTGCAT 100
 XX 817 GAATGACATTTGCTGTAGAGACCCCTTCAAGTGTGATCTGATCTCCAGAGCATTTGT 876
 DB 99 GAATGACATTTGCTGTAGAGACCCCTTCAAGTGTGATCTGATCTCCAGAGCATTTGT 40

XX 877 GGATGCTTACCCACCACTGATGATCTGCTGCACACT 915
 DB 39 GGATGCTTACCCACCACTGATGATCTGCTGCACACT 1

RESULT 10
 ID AA128873 standard; cDNA; 578 BP.
 XX AA128873;
 AC AA128873;
 XX 12-OCT-2001 (first entry)

XX Colon tumor related determined cDNA sequence for clone 31985.

XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 KW immunogenic; gene therapy; vaccine; colonic cancer; ss.
 XX Homo sapiens.

XX WO200149716-A2.
 PD 12-JUL-2001.

XX 29-DEC-2000; 2000MO-US35596.
 PF 30-DEC-1999; 99US-0476296.
 XX 10-JAN-2000; 2000US-0480321.
 PR 15-FEB-2000; 2000US-0504629.
 PR 06-MAR-2000; 2000US-0519444.
 PR 19-MAY-2000; 2000US-0575251.
 PR 29-JUN-2000; 2000US-0609448.
 PR 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIXA CORP.

XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
 PI King GE, Wang T, Jiang Y;
 DR WPI: 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -
 PS Claim 2: Page 249; 472pp; English.

XX The present invention describes colon tumor associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumor associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patient's own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
 CC given in the exemplification of the present invention.

XX Sequence 578 BP; 155 A; 139 C; 99 G; 185 T; 0 other;

Query Match 44.0%; Score 560.6; DB 22; Length 578;
Best Local Similarity 99.1%; Pred. No. 3.6e-149;
Matches 574; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 337 TTTGTCCTCCGATTAAGAGATATATCTGGAATATGTCCTGATCCCTGAGATTTGGGAAGA 396
DB 578 TTTGTCCTCCGATTAAGAGATATATCTGGAATATGTCCTGATCCCTGAGATTTGGGAAGA 519
QY 397 CCCAGGGGACAAATATAAACAATGCTGCTGTTGTGTGACAAATAGCAATGATGTTGTG 456
DB 518 CCCAGGGGACAAATATAAACAATGCTGCTGTTGTGTGACAAATAGCAATGATGTTGTG 459
QY 457 TGAATTTGGAAGCAAGATATGTCGAAGGTCGAATATATGCGCTGAAGTTTACAGCAT 516
DB 458 TGAATTTGGAAGCAAGATATGTCGAAGGTCGAATATAT- AATTGAAGTTTACAGCAT 400
QY 517 ATTGGCTATGATTTAGCAAGGGAAGCAAGCACTGGAAGTCAATTCATTAATGTCGATGA 576
DB 399 ATTGGCTATGATTTAGCAAGGGAAGCAAGCACTGGAAGTCAATTCATTAATGTCGATGA 340
QY 577 TCTGTATGTCAGCCCAATTAATTAATGATATGATGTCAACGGCTGAACCTGGCTACTT 636
DB 339 TCTGTATGTCAGCCCAATTAATTAATGATATGATGTCAACGGCTGAACCTGGCTACTT 280
QY 637 AGAAGCTACTGTCGACCTGTTAGAGGTATAGGTTCCCTGATGGAACACAGAAATGA 696
DB 279 AGAAGCTACTGTCGACCTGTTAGAGGTATAGGTTCCCTGATGGAACACAGAAATGA 220
QY 697 GTTTCGCTTTAATGACAGATTTAAAGATAGAGACTTTGCCATTAATTAATTAAGACAC 756
DB 219 GTTTCGCTTTAATGACAGATTTAAAGATAGAGACTTTGCCATTAATTAATTAAGACAC 160
QY 757 TCATGACCATTTGGAAGCATTAAGTACTAGAAACGATGGAAGCAATCACTTGAT 816
DB 159 TCATGACCATTTGGAAGCATTAAGTACTAGAAACGATGGAAGCAATCACTTGAT 100
QY 817 GAATACAACTTGTCTGTGAGACCCCTCAAGTGTGATGCTGTCGACAGCATGCT 876
DB 99 GAATACAACTTGTCTGTGAGACCCCTCAAGTGTGATGCTGTCGACAGCATGCT 40
QY 877 GGATGCTTTTACCACACCCCTGTGATGCTGCTGCACACT 915
DB 39 GGATGCTTTTACCACACCCCTGTGATGCTGCTGCACACT 1

RESULT 11
AA13542/c
ID AA13542 standard; DNA; 411 BP.
AC AA13542;
XX
XX 12-OCT-2001 (first entry)
DE Probe #3475 for gene expression analysis in human cervical cell sample.
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell.
KW cervical cancer; ss.
OS
OS Homo sapiens.
XX
XX
XX MO200157278-A2.
XX
XX
XX 09-AUG-2001.
XX
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488901/53.
XX
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX
PS Claim 25; SEQ ID NO 3475; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX
SQ Sequence: 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;

Query Match 30.4%; Score 387; DB 22; Length 411;
Best Local Similarity 96.4%; Pred. No. 5.8e-100;
Matches 396; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 273 CAAGGACCCCTTAACCCCTTAACCAAGATGGAAGAAAGAAACTGCTATGTTG 332
DB 411 CAAGGACCCCTTAACCCCTTAACCAAGATGGAAGAAAGAAACTGCTATGTTG 352
QY 333 CGAATTTGTCCTGATTAAGATATATCTGGAATATGTCCTGATGCTGATGCTG 392
DB 351 CGAATTTGTCCTGATTAAGATATATCTGGAATATGTCCTGATGCTGATGCTG 292
QY 393 AAGACCCGAGCAGATGATTAACATGCTGCTGTTGTGTGACAAATGCAATGATG 452
DB 291 AAGACCCGAGCAGATGATTAACATGCTGCTGTTGTGTGACAAATGCAATGATG 232
QY 453 TGTGTGAATTTGGAAGCAAGTATGTCGAAGGGAATTAATGCGCTGAAGTTGAG 512
DB 231 TGTGTGAATTTGGAAGCAAGTATGTCGAAGGGAATTAATGCGCTGAAGTTGAG 172
QY 513 GCATATTTGCTATGATTTGACGAAGGGGAACCGACTGGAAGTCAATTCATTAATG 572
DB 171 GCATATTTGCTATGATTTGACGAAGGGGAACCGACTGGAAGTCAATTCATTAATG 112
QY 573 ATGATCTGATGACAGCAATTAATGATATCAATGATGATCAACAGCTGAACCTGCT 632
DB 111 ATGATCTGATGACAGCAATTAATGATATCAATGATGATCAACAGCTGAACCTGCT 52
QY 633 ACTTAGAGCTACTGCTGATGCTGTTAGAAGTATTAAGTTCCTGATGAA 683
DB 51 ACTTAGAGCTACTGCTGATGCTGTTAGAAGTATTAAGTTCCTGATGAA 1

RESULT 12
AA134904/c
ID AA134904 standard; DNA; 411 BP.
AC AA134904;
XX
XX
XX 17-OCT-2001 (first entry)
XX
XX
XX Probe #3590 used to measure gene expression in human placenta sample.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.

```
XX OS Homo sapiens.
XX PN W0200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-48897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta
XX PS Claim 25; SEQ ID NO 3590; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;
Query Match 30.4%; Score 387; DB 22; Length 411;
Best Local Similarity 96.4%; Pred. No. 5.8e-100; Indels 0; Gaps 0;
Matches 396; Conservative 0; Mismatches 15;
OY 273 CAAAGGACCCCTTAAACCCCTTAAACAGATGTGAAAAAGGAAACCTCGCTATGTTG 332
DB 411 CAAAGGACCCCTTAAACCCCTTAAACAGATGTGAAAAAGGAAACCTCGCTATGTTG 352
OY 333 CGAATTTGTTCCCGTATTAAGATATATCTGGAATCTATGTCATCCCTGACACTGGG 392
DB 351 CGAATTTGTTCCCGTATTAAGATATATCTGGAATCTATGTCATCCCTGACACTGGG 292
OY 393 AAGACCCAGGACACATATGATTAACATGCTGCTGTGTGAGACATGACCAATTGATG 452
DB 291 AAGACCCAGGACACATATGATTAACATGCTGCTGTGTGAGACATGACCAATTGATG 232
OY 453 TGTGTGAATTTGGAAGCAAGGTATGTGCAAGAGTGAAATATTTGGCTGAAAGTTCTAG 512
DB 231 TGTGTGAATTTGGAAGCAAGGTATGTGCAAGAGTGAAATATTTGGCTGAAAGTTCTAG 172
OY 513 GCATATTTGCTATGATTTGCAAGAGGGAACCGACTGGAAGTCAATTCCTTAATGTGG 572
DB 171 GCATATTTGCTATGATTTGCAAGAGGGAACCGACTGGAAGTCAATTCCTTAATGTGG 112
OY 573 ATGATTCCTGATGACCAATTAATGATATCAATGATGTCAAAAGCTGGAACCTGGCT 632
DB 111 ATGATTCCTGATGACCAATTAATGATATCAATGATGTCAAAAGCTGGAACCTGGCT 52
OY 633 ACTTAAGACTACTGTGGACTGGTTAGAAAGTATTAAGTTCTCTGATGAA 683
DB 51 ACTTAAGACTACTGTGGACTGGTTAGAAAGTATTAAGTTCTCTGATGAA 1
RESULT 13
AA103431/C
ID AA103431 standard; DNA: 411 BP.
```

```
XX AC AA103431;
XX DT 09-OCT-2001 (first entry)
XX DE Probe #3422 used to measure gene expression in human breast sample.
XX KW Probe: human; breast disease; breast cancer; development disorder; ss;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN W0200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-476286/51.
XX PT Novel single exon nucleic acid probe used to measuring gene expression
XX PT in a human breast
XX PS Claim 25; SEQ ID NO 3422; 322pp; English.
XX CC The present invention relates to novel single exon nucleic acid probes.
XX CC The present sequence is one such probe. The probes are useful for
XX CC measuring human gene expression in a human breast sample, where the probe
XX CC hybridises at high stringency to a nucleic acid expressed in the human
XX CC breast. The probes are useful for predicting, diagnosing, grading,
XX CC staging, monitoring and prognosing diseases of the human breast,
XX CC particularly those diseases with polygenic aetiology. The diseases
XX CC include: breast cancer, disorders of development, inflammatory diseases
XX CC of the breast, fibrocystic changes, proliferative breast disease and
XX CC non-carcinoma tumours.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;
Query Match 30.4%; Score 387; DB 22; Length 411;
Best Local Similarity 96.4%; Pred. No. 5.8e-100; Indels 0; Gaps 0;
Matches 396; Conservative 0; Mismatches 15;
OY 273 CAAAGGACCCCTTAAACCCCTTAAACAGATGTGAAAAAGGAAACCTCGCTATGTTG 332
DB 411 CAAAGGACCCCTTAAACCCCTTAAACAGATGTGAAAAAGGAAACCTCGCTATGTTG 352
OY 333 CGAATTTGTTCCCGTATTAAGATATATCTGGAATCTATGTCATCCCTGACACTGGG 392
DB 351 CGAATTTGTTCCCGTATTAAGATATATCTGGAATCTATGTCATCCCTGACACTGGG 292
OY 393 AAGACCCAGGACACATATGATTAACATGCTGCTGTGTGAGACATGACCAATTGATG 452
DB 291 AAGACCCAGGACACATATGATTAACATGCTGCTGTGTGAGACATGACCAATTGATG 232
OY 453 TGTGTGAATTTGGAAGCAAGGTATGTGCAAGAGTGAATTAATTTGGCTGAAAGTTCTAG 512
DB 231 TGTGTGAATTTGGAAGCAAGGTATGTGCAAGAGTGAATTAATTTGGCTGAAAGTTCTAG 172
```


xx Claim 3, page 82, 89pp: English.

xx The invention provides novel human regulatory proteins (AAV28915-926),
xx designated HKRP, and their polynucleotides (AAV289285-296). The proteins
xx can be produced using standard recombinant technology. The expression of
xx HKRP is closely associated with cell proliferation and the polypeptides
xx and polynucleotides are useful in the diagnosis, treatment and
xx prevention of diseases associated with cell proliferation, particularly
xx immune responses and cancer. The protein or agonists may be administered
xx to treat or prevent a cancer such as adenocarcinoma, leukemia, lymphoma,
xx melanoma, myeloma, sarcoma, and tetracarcinoma. Such cancers include,
xx but are not limited to, cancers of the adrenal gland, bladder, bone,
xx brain, cervix, breast, gall bladder, ganglia, ovary and pancreas. Where
xx HKRP is promoting leukocyte activity or proliferation, antagonists which
xx decrease the activity of HKRP are administered. Such responses may be
xx associated with disorders such as Addison's disease, adult respiratory
xx distress syndrome, AIDS, allergies, anemia, asthma, atherosclerosis,
xx bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic
xx dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic
xx gastritis, gout, glomerulonephritis, Grave's disease, hypersinophilia,
xx irritable bowel syndrome, lupus erythematosus, multiple sclerosis,
xx myasthenia gravis, myocardial or pericardial infarction, osteoarthritis,
xx osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis,
xx scleroderma, Sjogren's syndrome, and autoimmune thyroiditis.
xx complications of cancer, hemodialysis, extracorporeal circulation; viral,
xx bacterial, fungal, parasitic, protozoan, and helminthic infections; and
xx trauma. The HKRP polynucleotide may be used for diagnosis of these
xx conditions, and as a source of primers and probes.

xx Sequence 1198 BP; 377 A; 227 C; 266 G; 328 T; 0 other;

xx Query Match 29.5%; Score 375.8; DB 20; Length 1198;
xx Best Local Similarity 69.6%; Pred. No. 1.6e-96;
xx Matches 509; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

xx 161 AAAGGACATATATATCCATTCATGATATTCATTTATGACAGATGTTT 220
xx 227 aaagaggaataatgcattccataagaagacagaaatgatgataatgcatt 296
xx 221 CACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 280
xx 297 aatgatgatgataatgcattccataagaagacagaaatgatgataatgcatt 356
xx 281 CCTTAAACCCATTAACAGATGATGATGATGATGATGATGATGATGATG 340
xx 357 ccaatgaatccatcaataaataatgaatgaatgaatgaatgaatgaatga 416
xx 341 TTCCCGTATTAAGATATATGATGATGATGATGATGATGATGATGATGAT 400
xx 417 ttcccttaagaagatataatgaatgaatgaatgaatgaatgaatgaatga 476
xx 401 GGGCACAATGATTAACATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTG 460
xx 477 catgaataaagaataagaacacgccttggaataatgaatgaatgaatga 536
xx 461 ATTGAAGCAAGGATGATGATGATGATGATGATGATGATGATGATGATG 520
xx 537 atagatcaagaatctctctctctctctctctctctctctctctctctct 596
xx 521 GCTATGATTAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 580
xx 597 gctctatgatgaagatgaagatgaagatgaagatgaagatgaagatgaag 656
xx 581 GATGACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 640
xx 657 gaagcccaagatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 716
xx 641 GCTACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 700
xx 717 gtactcttaattgattatataatgaatgaatgaatgaatgaatgaatgaat 776

xx 701 GCGTTAATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 760
xx 777 gcttcaatgaagaaatcaaaaacagccttctctgaatgaatgaatgaatga 836
xx 761 GACCATTTGAGAGGATTAATTAATTAATTAATTAATTAATTAATTAAT 820
xx 837 caatgttgaagacatctctatgaagacatgaatgaagacatgaatgaag 896
xx 821 ACAACCTTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 880
xx 897 gtgcagatctcgaatgaatgaatgaatgaatgaatgaatgaatgaatga 956
xx 881 GCTTTACACAC 891
xx 957 tggatctatc 967

Search completed: March 29, 2002, 19:55:03
Job time: 1442 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 18:52:37 ; Search time 1849.3 Seconds

(without alignments)
11373.981 Million cell updates/sec

Title: US-09-415-540-2

Perfect score: 1275

Sequence: 1 CAAGAGCTTGGGGCTCTCT.....GAATTAATTAATTTGCTG 1275

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vl: *
30: em_htgo_hum: *
31: em_htgo_inv: *
32: em_htgo_rod: *
33: em_htg_hum: *
34: em_htg_inv: *
35: em_htg_rod: *
36: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1274	99.9	1275	6	AR061907 Sequence
2	1274	99.9	1275	6	AR084921 Sequence
3	1266.8	99.4	1303	6	AX015063 Sequence
4	1265.8	99.3	1301	9	BC001022 Homo sapi
5	1264	99.1	1282	9	AF119665 Homo sapi
6	1228.8	96.4	1243	9	AF154065 Homo sapi
7	1217	95.5	1239	9	AB026723 Homo sapi
8	1185.4	93.0	1204	9	AF217186 Homo sapi
9	1179.4	92.5	1200	6	AX018061 Sequence
10	1123.8	88.1	167012	2	AC008429 Homo sapi
11	1123.8	88.1	213025	2	AC021159 Homo sapi
12	1079.2	84.6	155645	2	AC011012 Homo sapi
13	917.2	71.9	1266	4	BOV026PP
14	841.2	66.0	846	9	AF108211
15	835.8	65.6	1243	10	BC010468
16	799.2	62.7	856	6	AX014862 Sequence
17	560.6	44.0	578	6	AX192855 Sequence
18	546.4	42.9	70575	2	AC025103 Homo sapi
19	391.2	30.7	1088	5	AF174582 Torpedo m
20	380.2	30.6	70575	2	AC025103 Homo sapi
21	375.8	29.5	1202	6	AX011614 Sequence
22	375.8	29.5	1214	9	AF217187 Homo sapi
23	375.8	29.5	1706	6	AX000466 Homo sapi
24	368.8	28.9	1188	6	AX086197 Sequence
25	368.8	28.9	1188	9	HS0801625
26	350	27.5	165756	2	AL355138 Homo sapi
27	350	27.5	215049	2	AC067749 Homo sapi
28	339.2	26.6	1225	10	BC011417 Mus muscu
29	238.2	18.7	1115	9	AB026722 Homo sapi
30	236.6	18.6	938	11	CNS06P1E
31	235.4	18.5	873	3	AF085601
32	229.6	18.0	1612	8	SCPPAG
33	229.6	18.0	2411	8	SCYBR011C
34	226.6	17.8	2887	8	KLTRPLOC
35	223	17.5	2326	6	PETPR1877
36	221.4	17.4	184180	9	HS1042K10
37	215	16.9	2759	8	ZBA309279
38	212.6	16.7	720	8	CNS01DAS
39	208	16.3	37855	3	CEC47E12
40	202.2	15.9	357	6	AX014834 Sequence
41	201.2	15.8	636	8	CNS01A30
42	191.8	15.0	1366	8	SPPAG
43	191.8	15.0	38391	8	SPAC23C11
44	190.8	15.0	911	8	ATH252210
45	190.8	15.0	1161	8	AY045945 Arabidops

ALIGNMENTS

RESULT	1	PAT	29-SEP-1999
AR061907	AR061907	1275 bp	DNA
LOCUS	Sequence 2 from patent US 5843665.		
DEFINITION	AR061907		
ACCESSION	AR061907.1	GI:5989598	
VERSION			
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1275)		
AUTHORS	Hawkins, P.R. and Hillman, J.L.		
TITLE	Human pyrophosphatase		
JOURNAL	Patent: US 5843665-A 2 01-DEC-1998;		
FEATURES	Location/Qualifiers		
source	1..1275		
BASE COUNT	394 a 235 c 294 g 351 t 1 others		
ORIGIN			

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	ORIGIN	BASE COUNT	Query Match	Best Local Similarity	Matches 1275; Conservative	99.9%; Score 1274; DB 6; length 1275; Pred. No. 1.3e-284;	0; Mismatches	0; Indels	0; Gaps
QY	1021	TAAAGTAGAGCTTTTCAAGAGCTTTAAATTTGTAGAACCTACTACTAAGTAATTC	1080																		
DB	1021	TAAAGTAGAGCTTTTCAAGAGCTTTAAATTTGTAGAACCTACTACTAAGTAATTC	1080																		
QY	1081	TGCGTGTACTAATCCAAATATPACTCAGAAATGTTATCCATTAAGCATTTTTCATATCTCA	1140																		
DB	1081	TGCGTGTACTAATCCAAATATPACTCAGAAATGTTATCCATTAAGCATTTTTCATATCTCA	1140																		
QY	1141	ACTAGATAACTTTTACGACATGCTTAAATATCAAGCAGTTGCTATTTTGGAAAGTCACCT	1200																		
DB	1141	ACTAGATAACTTTTACGACATGCTTAAATATCAAGCAGTTGCTATTTTGGAAAGTCACCT	1200																		
QY	1201	GTGAATAGATGTGCAAGGGGAGCACAATTTGGATGTATGTATGTACCATATGTTAGAAAT	1260																		
DB	1201	GTGAATAGATGTGCAAGGGGAGCACAATTTGGATGTATGTATGTATGTACCATATGTTAGAAAT	1260																		
QY	1261	AAAATATATTTTGGCTG 1275																			
DB	1261	AAAATATATTTTGGCTG 1275																			
QY	1261	AAAATATATTTTGGCTG 1275																			
DB	1261	AAAATATATTTTGGCTG 1275																			
QY	1261	AAAATATATTTTGGCTG 1275																			
DB	1261	AAAATATATTTTGGCTG 1275																			
QY	1261	AAAATATATTTTGGCTG 1275																			
DB	1261	AAAATATATTTTGGCTG 1275																			
QY	1261	AAAATATATTTTGGCTG 1275																			
DB	1261	AAAATATATTTTGGCTG 1275																			
QY	1261	AAAATATATTTTGGCTG 1275																			
DB	1261	AAAATATATTTTGGCTG 1275																			
QY	1261	AAAATATATTTTGGCTG 1275																			
DB	1261	AAAATATATTTTGGCTG 1275																			
QY	1261	AAAATATATTTTGGCTG 1275																			
DB	1261	AAAATATATTTTGGCTG 1275																			
QY	1261	AAAATATATTTTGGCTG 1275																			
DB	1261	AAAATATATTTTGGCTG 1275																			
QY	1261	AAAATATATTTTGGCTG 1275																			

```

QY 421 TGGCTGTTGTGTCACAAATGACCAATGATGTGTGAATTTGGAAGCAAGTAATGTC 480
    |||||
Db 421 TGGCTGTTGTGTCACAAATGACCAATGATGTGTGAATTTGGAAGCAAGTAATGTC 480
QY 481 AAGAGGTGAATTAATGCGCGTGAAGTTCTAGCAATTTGCTATGATTTGACGAAGGGA 540
    |||||
Db 481 AAGAGGTGAATTAATGCGCGTGAAGTTCTAGCAATTTGCTATGATTTGACGAAGGGA 540
QY 541 AAGCAGCTGGAAGTATGTCATTAATGATGATGCTGATGAGCAATTAATTAATGA 600
    |||||
Db 541 AAGCAGCTGGAAGTATGTCATTAATGATGATGCTGATGAGCAATTAATTAATGA 600
QY 601 TATCAATGATGTCACAAAGCGCTGAACCTGCTACTTAGAAGTACTGTGAGCTGGTTAG 660
    |||||
Db 601 TATCAATGATGTCACAAAGCGCTGAACCTGCTACTTAGAAGTACTGTGAGCTGGTTAG 660
QY 661 AAGGTATTAAGTTCTGATGAGAAAACAGAAAATGATTTGCGTTTAATGCAAAATTTAA 720
    |||||
Db 661 AAGGTATTAAGTTCTGATGAGAAAACAGAAAATGATTTGCGTTTAATGCAAAATTTAA 720
QY 721 AATATAGGACTTGGCATGATATTAATTAAGCACTGATGACCAATGGAAGCAATTAAT 780
    |||||
Db 721 AATATAGGACTTGGCATGATATTAATTAAGCACTGATGACCAATGGAAGCAATTAAT 780
QY 781 GACTAAGAAAGCAATGAGAAAAGCAATGATGATGATGATGATGATGATGATGATGAT 840
    |||||
Db 781 GACTAAGAAAGCAATGAGAAAAGCAATGATGATGATGATGATGATGATGATGATGAT 840
QY 841 CTTCAGTGTGATGCTGATGCTGACGAGCAATGATGATGATGATGATGATGATGATGAT 900
    |||||
Db 841 CTTCAGTGTGATGCTGATGCTGACGAGCAATGATGATGATGATGATGATGATGATGAT 900
QY 901 ATCTGCTGACAGTACAGCAAGAGCTGATGATGATGATGATGATGATGATGATGATGAT 960
    |||||
Db 901 ATCTGCTGACAGTACAGCAAGAGCTGATGATGATGATGATGATGATGATGATGATGAT 960
QY 961 AATTTCTGTAATGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
    |||||
Db 961 AATTTCTGTAATGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 1021 TAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
    |||||
Db 1021 TAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
QY 1081 TGGTGTGATTAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
    |||||
Db 1081 TGGTGTGATTAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 ACTAAGATTAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
    |||||
Db 1141 ACTAAGATTAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 GTGAATAGATGTCAGAGGAGGACATATGATGATGATGATGATGATGATGATGATGAT 1260
    |||||
Db 1201 GTGAATAGATGTCAGAGGAGGACATATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 AAAATTTATTTGCTG 1275
    |||||
Db 1261 AAAATTTATTTGCTG 1275

```

```

RESULT 3
AX015063 1303 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 272 from Patent W09953040.
ACCESSION AX015063
VERSION AX015063.1 GI:10041202
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

```

```

REFERENCE 1 (bases 1 to 1303)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
TITLE Human nucleic acid sequences from ovarian tumour tissue
JOURNAL Patent: WO 9953040-A 272 21-OCT-1999; DAHL EDGAR (DE); HINZMANN
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); PILARSKY CHRISTIAN (DE)
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
source 1..1303
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 410 a 240 c 300 g 353 t
ORIGIN
Query Match: 99.4%; Score 1266.8; DB 6; Length 1303;
Best Local Similarity 99.8%; Pred. No. 6e-283;
Matches 1268; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 AGTTGGGGCTCTCCCTTGCAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 64
    |||||
Db 11 AGCGGTGGGCTCTCTCTTGTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 70
QY 65 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 124
    |||||
Db 71 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 130
QY 125 TTCTCCCTGAGTACCGATCTTCTCAAAAATGAGAAAGCAATATATATCTCCATT 184
    |||||
Db 131 TTCTCCCTGAGTACCGATCTTCTCAAAAATGAGAAAGCAATATATATCTCCATT 190
QY 185 CATGATATTCATTTATGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 244
    |||||
Db 191 CATGATATTCATTTATGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 250
QY 245 TGGTCTAATGCAAAATGAGATGCTCAAAAGCAACCTTTAAACCTATTAAACAGAT 304
    |||||
Db 251 TGGTCTAATGCAAAATGAGATGCTCAAAAGCAACCTTTAAACCTATTAAACAGAT 310
QY 305 GTGAAAAAAGAAACCTGCTATGTTGCAATTTGTTCCGTATTAAGATATATCTGG 364
    |||||
Db 311 GTGAAAAAAGAAACCTGCTATGTTGCAATTTGTTCCGTATTAAGATATATCTGG 370
QY 365 AACTATGTCGATCCCTGAGCTGGGAGGACCGGACCAATGATTAACATCTGGC 424
    |||||
Db 371 AACTATGTCGATCCCTGAGCTGGGAGGACCGGACCAATGATTAACATCTGGC 430
QY 425 TGGTGTGATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 484
    |||||
Db 431 TGGTGTGATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 490
QY 485 GGTGAATTAATTTGGCGTGAAGTTCTAGGCAATTTGGCTATGATGACGAAGGGAACC 544
    |||||
Db 491 GGTGAATTAATTTGGCGTGAAGTTCTAGGCAATTTGGCTATGATGACGAAGGGAACC 550
QY 545 GACTGGAAGTATTTGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 604
    |||||
Db 551 GACTGGAAGTATTTGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 610
QY 605 AATGATGTCAAAGCGCTGAACCTGCTAGAGAGTACTGAGTACTGAGTACTGAGTACTGAGT 664
    |||||
Db 611 AATGATGTCAAAGCGCTGAACCTGCTAGAGAGTACTGAGTACTGAGTACTGAGTACTGAGT 670
QY 665 TATTAAGTTCTGATGAGAAACAGAAATGATGATGATGATGATGATGATGATGATGATGAT 724
    |||||
Db 671 TATTAAGTTCTGATGAGAAACAGAAATGATGATGATGATGATGATGATGATGATGATGAT 730
QY 725 AAGGACTTTGCCATGATATTTATTAAGCACTGATGATGATGATGATGATGATGATGATGAT 784
    |||||
Db 731 AAGGACTTTGCCATGATATTTATTAAGCACTGATGATGATGATGATGATGATGATGATGAT 790
QY 785 AAGAAAGCAATGAGAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 844

```


RESULT	5			
LOCUS	AF119665			
DEFINITION	AF119665	1282 bp	mRNA	PRI
ACCESSION	AF119665			13-DEC-1999
VERSION	AF119665			complete cds.
KEYWORDS	AF119665.1	GI:6563255		
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
JOURNAL	1 (bases 1 to 1282)			
REFERENCE	Gu,Y., Peng,Y., Li,Y., Fu,S., Gu,J., Zhang,L., Jiang,C., Yu,Y.,			
AUTHORS	Han,Z., Wang,Y., Chen,Z., and Fu,G.			
JOURNAL	A novel gene expressed in the human adrenal gland			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 1282)			
JOURNAL	Huang,C., Zhang,C., Wu,T., Peng,Y., Gu,Y., Zhang,L., Jiang,C.,			
REFERENCE	Li,Y., Han,Z., Wang,Y., Chen,Z., and Fu,G.			
AUTHORS	Direct Submission			
JOURNAL	Submitted (12-JAN-1999) Chinese National Human Genome Center at			
REFERENCE	Shanghai, Guo Shoujing Rd., 351, Zhangjiang Hi-Tech Park, Pudong			

FEATURES	Shanghai 201203, P. R. China
SOURCE	Location/Qualifiers
	1. 1282
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/tissue_type="adrenal gland"
	78. 947
CDS	/note="IOPPP"
	/evidence-not_experimental
	/product="inorganic pyrophosphatase"
	/protein_id="AA171222.1"
	/db_xref="GI:6563256"
	/translation="MSGSFSEERAAPFSLERYFLNKEGQVISPFIIDPIYDKDVE HMYEPRWSNKMELATKDPINPIKDYKKGILRYANLFPKGIWNGALPQTE DPGNKHTKCCGGNDP IDVCEIGSKAGELIGVLCILIMIDEGEDMIVILN VDPDPAANTYNDINDVYKRLKPGYLEATYDWMFRKRVKPGKRENFARNAEKKDFPID LKSTHDMKALVTYTKTNGKISCMNTTLESFPKCDPDAIRAIYDALPPCSACTV PTVDKMFHOKN"
BASE COUNT	409 a 234 c 290 g 349 t
ORIGIN	
Query Match	99.18; Score 1264; DB 9; Length 1282;
Best Local Similarity	100.0%; Prid. No. 2.7e-282;
Matches 1264; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	12 GGGGCTCTCTCTGTCATGTCGGCGCGGTCGGGCTGGTGGCTGTGGCAGCGCGGC 71
DB	1 GGGCTCTCTCTGTCATGTCGGCGCGGTCGGGCTGGTGGCTGTGGCAGCGCGGC 60
OY	72 GGCAGACTCCGGCAGCTATGAGCGGCTTCACACCGAGAGACGCGCGCCCTTCTCC 131
DB	61 GGCAGACTCCGGCAGCTATGAGCGGCTTCACACCGAGAGACGCGCGCCCTTCTCC 120
OY	132 TGGAGTACCGAGCTCTCTCTCAAAATGGAAGGACATATATATCTCATTTATGATA 191
DB	121 TGGAGTACCGAGCTCTCTCTCAAAATGGAAGGACATATATATCTCATTTATGATA 180
OY	192 TTCGAATTTATGAGATTAAGATGTGTTTCACATGTGTAGTGAAGTACACAGCTGCTTA 251
DB	181 TTCGAATTTATGAGATTAAGATGTGTTTCACATGTGTAGTGAAGTACACAGCTGCTTA 240
OY	252 ATGCAAAATGAGATGCTCTACAAAGACCCCTTTAAACCTATTAAACAGATGTGAAA 311
DB	241 ATGCAAAATGAGATGCTCTACAAAGACCCCTTTAAACCTATTAAACAGATGTGAAA 300
OY	312 AAGGAAAATCTCGCTATGTTCGGAATTTGTTCCGCTATTAAGGATATATCTGGAATATG 371
DB	301 AAGGAAAATCTCGCTATGTTCGGAATTTGTTCCGCTATTAAGGATATATCTGGAATATG 360
OY	372 GTGGCATCCCTCAGACTTGGAGAGACCCAGGAGCAATGATTAACATATCTGGCGTTGTG 431
DB	361 GTGGCATCCCTCAGACTTGGAGAGACCCAGGAGCAATGATTAACATATCTGGCGTTGTG 420
OY	432 GTGACAAATGACCAATTGATGTGTGTAATTTGGAAGCAAGGTATGTGCAAGAGGTGAAA 491
DB	421 GTGACAAATGACCAATTGATGTGTGTAATTTGGAAGCAAGGTATGTGCAAGAGGTGAAA 480
OY	492 TAATTTGGCGTGAAGTTCTTAGGCATATTGGCTATGATTTGACGAAGGGGAAAACCGACTGGA 551
DB	481 TAATTTGGCGTGAAGTTCTTAGGCATATTGGCTATGATTTGACGAAGGGGAAAACCGACTGGA 540
OY	552 AAGTCATTGCAATTAATGTGATGATCCTGTATGAGCCAAATTAATGATATCAATGATG 611
DB	541 AAGTCATTGCAATTAATGTGATGATCCTGTATGAGCCAAATTAATGATATCAATGATG 600
OY	612 TCAAAACGGCTGAACCTGGCTACTTAGAAGCTACTGTGAGCTGTTTGAAGGTATTAAGG 671
DB	601 TCAAAACGGCTGAACCTGGCTACTTAGAAGCTACTGTGAGCTGTTTGAAGGTATTAAGG 660
OY	672 TTCCGATGGAAGAACAGAAAATGAGTTGGCTTTAATGCAATTTTAAATGATTAAGGACT 731
DB	661 TTCCGATGGAAGAACAGAAAATGAGTTGGCTTTAATGCAATTTTAAATGATTAAGGACT 650

Db 661 TTCCATGATGAAACCAAGAAATAGTTCGCTTAAATGCAAAATTAAGATAAGCACT 720
 Oy 732 TTGGCATGATATTTTAAAGCACTCATGACCATTTGAAAGCACTTAGTACATAAGAAA 791
 Db 721 TTGGCATGATATTTTAAAGCACTCATGACCATTTGAAAGCACTTAGTACATAAGAAA 780
 Oy 792 CGAATGAAAGAAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATG 851
 Db 781 CGAATGAAAGAAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATG 840
 Oy 852 ATCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 911
 Db 841 ATCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 Oy 912 CAGTACCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 971
 Db 901 CAGTACCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
 Oy 972 GAATACCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1031
 Db 961 GAATACCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
 Oy 1032 GCTTTCAAGCTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1091
 Db 1021 GCTTTCAAGCTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1080
 Oy 1092 ATCCATATATCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1151
 Db 1081 ATCCATATATCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 Oy 1152 TTTTACCAATGCTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1211
 Db 1141 TTTTACCAATGCTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 Oy 1212 TGCAGGAGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1271
 Db 1201 TGCAGGAGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
 Oy 1272 GCTG 1275
 Db 1261 GCTG 1264

RESULT 6
 AF154065 1243 bp mRNA PRI 09-NOV-1999
 DEFINITION Homo sapiens inorganic pyrophosphatase (pp) mRNA, complete cds.
 ACCESSION AF154065
 VERSION AF154065.1 GI:4960207
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Fairchild T.A. and Patejunas G.
 TITLE Cloning and expression profile of human inorganic pyrophosphatase
 JOURNAL Biochim. Biophys. Acta 1447 (2-3), 133-136 (1999)
 MEDLINE 20011271
 REFERENCE 2 (bases 1 to 1243)
 AUTHORS Patejunas G.
 TITLE Direct Submission
 JOURNAL Submitted (25-MAY-1999) Molecular Cardiology Laboratory, Evanston
 Northwestern Healthcare Research Institute, 2650 Ridge Avenue,
 Evanston, IL 60201, USA

FEATURES
 source
 1. 1243
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /map="10q21.3"
 /tissue_type="heart"
 1. 1243

CDS
 /gene="pp"
 46..915
 /CDS_number="3.6.1.1"
 /function="catalyzes hydrolysis of pyrophosphate"
 /codon_start=1
 /product="inorganic pyrophosphatase"
 /protein_id="A03463.1"
 /db_xref="GI:4960208"
 /translation="MSGFSTEREAPFSLERVFLKNEKQYISPPHDIPIYADKDF
 HMYEVRMSNMKEIATKDPINPIKODYKRLRYANLFPYKGYIMNGATIPOTWE
 DRHNKRNKGGCGNDPIDVCEIGSVARGLIIGVKYGLIIMDEBETDKRLAIN
 VDDPDANINDINDYRLKPYGLAETVDMFRYKVPDKEPEPEFNFKDFDAID
 IKTSDHMKALVTYKTKNGKGISCAWTTISSEPFKCPDDPAAIYDALPPESACTV
 PFDVDMFPHOKN"

BASE COUNT 400 a 224 c 279 g 340 t
 ORIGIN

Query Match 96.4%; Score 1228.8; DB 9; Length 1243;
 Best Local Similarity 99.8%; Pred. No. 3.8e-274;
 Matches 1230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 44 GGCGTGTGCTCTGTGACAGCGCGCGGACGAGCATCGGCACTATGAGCGCTTACG 103
 Db 1 GGCGTGTGCTCTGTGACAGCGCGCGGACGAGCATCGGCACTATGAGCGCTTACG 60
 Oy 104 ACCGAGGAGCGCGCGCGCGCTTCCCTGAGTACCGAGTCTTCTCAAAAATGAGAAA 163
 Db 61 ACCGAGGAGCGCGCGCGCGCTTCCCTGAGTACCGAGTCTTCTCAAAAATGAGAAA 120
 Oy 164 GGACATATATATCTCATCTTTCATGATATTTCAATTTATGACAGTAAGATGCTTTCAC 223
 Db 121 GGACATATATATCTCATCTTTCATGATATTTCAATTTATGACAGTAAGATGCTTTCAC 180
 Oy 224 ATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 283
 Db 181 ATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
 Oy 284 TTTAAACCTATTTAAACAAGATGTAAGAAAAGAAAACCTTCCCTATGTCGAATTTGTC 343
 Db 241 TTTAAACCTATTTAAACAAGATGTAAGAAAAGAAAACCTTCCCTATGTCGAATTTGTC 300
 Oy 344 CCGTATTAAGATATATCTGGAACATGATGATGATGATGATGATGATGATGATGATGAT 403
 Db 301 CCGTATTAAGATATATCTGGAACATGATGATGATGATGATGATGATGATGATGATGAT 360
 Oy 404 CACATGATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 463
 Db 361 CACATGATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 Oy 464 GGAAGCAAGTATGTCAGAGGTGAATTAATTTGGCTGGAAGTCTTGGCATATTGGCT 523
 Db 421 GGAAGCAAGTATGTCAGAGGTGAATTAATTTGGCTGGAAGTCTTGGCATATTGGCT 480
 Oy 524 ATGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 583
 Db 481 ATGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
 Oy 584 GCAGCAATTAATGATATCAATGATGATGATGATGATGATGATGATGATGATGATGATG 643
 Db 541 GCAGCAATTAATGATATCAATGATGATGATGATGATGATGATGATGATGATGATGATG 600
 Oy 644 ACTGTGACTGTTTGAAGATTAAGTTCCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 703
 Db 601 ACTGTGACTGTTTGAAGATTAAGTTCCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
 Oy 704 TTTAATGACAGATTTAAGATTAAGATTTGATGATGATGATGATGATGATGATGATGATG 763
 Db 661 TTTAATGACAGATTTAAGATTAAGATTTGATGATGATGATGATGATGATGATGATGATG 720
 Oy 764 CATTGGAAGCATTAAGTGAAGAAAAGCAATGAGAAAAGATGATGATGATGATGATGATG 823

Db 721 CATTGGAAAGCATAGTACTAGAGAAAGCAATGGAAAAAGCAATGATTCATGATGATA 780
 Oy 824 ACTTGTCTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAGCCATTGTGATGCT 883
 Db 781 ACTTGTCTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAGCCATTGTGATGCT 840
 Oy 884 TTACACCAACCCCTGTGATCTGCTGACAGACAGACAGAGAGAGTAAAGTGGTTCAT 943
 Db 841 TTACACCAACCCCTGTGATCTGCTGACAGACAGACAGAGAGAGTAAAGTGGTTCAT 900
 Oy 944 CACCAAGAAAACATAAGATTTCTGTGAAATACAGAGTATGCTACATGCTGTCA 1003
 Db 901 CACCAAGAAAACATAAGATTTCTGTGAAATACAGAGTATGCTACATGCTGTCA 960
 Oy 1004 TCTGATGATAGAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1063
 Db 961 TCTGATGATAGAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1020
 Oy 1064 CTAACTAAAGTAAATTTCTGCTGCTGACATATCCATATCTCAGATGTTATCTATAA 1123
 Db 1021 CTAACTAAAGTAAATTTCTGCTGCTGACATATCCATATCTCAGATGTTATCTATAA 1080
 Oy 1124 GCATTTTCATATCTCACTAAGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1183
 Db 1081 GCATTTTCATATCTCACTAAGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1140
 Oy 1184 TCATTTGAGTACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1243
 Db 1141 TCATTTGAGTACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
 Oy 1244 ACCATATGTTAGAAATAAATTTATTTGCTG 1275
 Db 1201 ACCATATGTTAGAAATAAATTTATTTGCTG 1232

RESULT 7
 AB026723 LOCUS AB026723 1239 bp mRNA PRI 29-SEP-1999
 DEFINITION Homo sapiens S1D6-8061 mRNA for pyrophosphatase, complete cds.
 ACCESSION AB026723.1 GI:5931601
 VERSION AB026723.1 GI:5931601
 KEYWORDS pyrophosphatase.
 SOURCE Homo sapiens tissue_c11b:brain cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 1239)
 AUTHORS Saito, T., Hattori, A. and Miyajima, N.
 TITLE Putative inorganic pyrophosphatase.
 JOURNAL Published Only in Database (1999) In press
 REFERENCE 2 (bases 1 to 1239)
 AUTHORS Saito, T., Hattori, A. and Miyajima, N.
 TITLE Direct Submission
 JOURNAL Submitted (26-APR-1999) to the DDBJ/EMBL/GenBank databases.
 Toshiyuki Saito, National Institute of Radiological Sciences,
 Genome Research Group; Anagawa 4-9-1, Inage-ku, Chiba 263-8555,
 Japan (E-mail: t.saito@nirs.go.jp, Tel: 81-43-206-3135,
 Fax: 81-43-251-9818)

FEATURES

source location/Qualifiers
 1. 1239
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /tissue_11b="brain"
 31. 900
 /gene="S1D6-8061"
 31. 900
 /gene="S1D6-8061"
 /codon_start=1
 /product="pyrophosphatase"
 /protein_id="BA084702.1"
 /db_xref="GI:5931602"
 /translation="MSGFSTEERAPFSLFLEYRVFLKNEKGOYISPFHDIPYADKDYV"

HAVEPEPMSNAKKEIATDLPNIPKODPKKGLRYVANLPFYKY IINVCALPOTHE
 DRGNHDKHTGCCGNDPIDVCEIGSKVAREIIGVKLGLILAMDEETPMKVAIN
 VDDPDANYNINIDVYKRLKPGYLETVDYEMFRYKYPDCKPERENAFNFDKDAID
 IIKSTHDMKALVTKINGKISLKNWTLISPSFKDPAARAIIVDALPPCESACTV
 PTDVDKMHKKN"
 BASE COUNT 407 a 220 c 273 g 339 t
 ORIGIN

Query Match 95.5%; Score 1217; DB 9; Length 1239;
 Best Local Similarity 100.0%; Pred. No. 2e-271;
 Matches 1217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 59 TGGCAGCGGGGGGGGAGAGACTCCGGCACTATAGAGGGGCTTACAGACCGAGAGCGCGCC 118
 Db 1 TGGCAGCGGGGGGAGAGACTCCGGCACTATAGAGGGGCTTACAGACCGAGAGCGCGCC 60
 Oy 119 GCGCCCTTCTCCCTGGAGTACCGAGTCTTCTCAAAAATGAGAAAGACATATATATCT 178
 Db 61 GCGCCCTTCTCCCTGGAGTACCGAGTCTTCTCAAAAATGAGAAAGACATATATATCT 120
 Oy 179 CCATTTTCATGATTTTCAATTTATGACATTAAGATGTGTTTACATGCTAGTGAAGTA 238
 Db 121 CCATTTTCATGATTTTCAATTTATGACATTAAGATGTGTTTACATGCTAGTGAAGTA 180
 Oy 239 CCACGCTGCTTAATGCAAAAATGAGATGCTACAAAAGACCTTTAAACCTATTAA 298
 Db 181 CCACGCTGCTTAATGCAAAAATGAGATGCTACAAAAGACCTTTAAACCTATTAA 240
 Oy 299 CAAGATGTGAAAAAGAAAGAACTGCTATGTTGGCAATTTTCCGTAATAAGATAT 358
 Db 241 CAAGATGTGAAAAAGAAAGAACTGCTATGTTGGCAATTTTCCGTAATAAGATAT 300
 Oy 359 ATCTGGAATATGTGTCATCCCTCAGACTTGGGAAGACCCAGGCGCAATGATAACAT 418
 Db 301 ATCTGGAATATGTGTCATCCCTCAGACTTGGGAAGACCCAGGCGCAATGATAACAT 360
 Oy 419 ACTGGCTGTGTGGTGCATCAATGACCAATGATGATGATGATGATGATGATGATGATG 478
 Db 361 ACTGGCTGTGTGGTGCATCAATGATGATGATGATGATGATGATGATGATGATGATG 420
 Oy 479 GCAAGAGTGAATATATTGGCTGAAAGTCTTAGGCATATTTGGCTATGATGATGATGATG 538
 Db 421 GCAAGAGTGAATATATTGGCTGAAAGTCTTAGGCATATTTGGCTATGATGATGATGATG 480
 Oy 539 GAAACCGACTGGAAGTCTATTTGCCATTAATGTGATGATGATGATGATGATGATGATG 598
 Db 481 GAAACCGACTGGAAGTCTATTTGCCATTAATGTGATGATGATGATGATGATGATGATG 540
 Oy 599 GATATCAATGATGTCAAAAGCGGTGAACCTGCTACTTGAAGCTCTGTGACTGGTT 658
 Db 541 GATATCAATGATGTCAAAAGCGGTGAACCTGCTACTTGAAGCTCTGTGACTGGTT 600
 Oy 659 AGAAGGTATAGCTTCTGATGATGAAAAAGCAAAATGATTTGCTTAAATGACAAATTT 718
 Db 601 AGAAGGTATAGCTTCTGATGATGAAAAAGCAAAATGATTTGCTTAAATGACAAATTT 660
 Oy 719 AAAGATAGGACTTTGCCATTTGATATTATTAAGACATGATGATGATGATGATGATGAT 778
 Db 661 AAAGATAGGACTTTGCCATTTGATATTATTAAGACATGATGATGATGATGATGATGAT 720
 Oy 779 GTGACTAAGAAAAGCAATGGAAGAAAGCAATGATGATGATGATGATGATGATGATGATG 838
 Db 721 GTGACTAAGAAAAGCAATGGAAGAAAGCAATGATGATGATGATGATGATGATGATGATG 780
 Oy 839 CCCTTCAAGTGTGATCCTGATGCTGACAGAGCATTTGATGATGATGATGATGATGATG 898
 Db 781 CCCTTCAAGTGTGATCCTGATGCTGACAGAGCATTTGATGATGATGATGATGATGATGAT 840
 Oy 899 GAATGCTGCTGACAGTACCAACAGAGCTGATTAAGTGTTCATATGACAGAAAACTAA 958
 Db 841 GAATGCTGCTGACAGTACCAACAGAGCTGATTAAGTGTTCATATGACAGAAAACTAA 900

Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 303711
Center clone name: CIT-HSPC_308K20

Summary Statistics

Consensus quality: 156314 bases at least Q40
Consensus quality: 158094 bases at least Q30
Consensus quality: 158732 bases at least Q20
Estimated insert size: 162000; pulse field gel estimation
Estimated insert size: 166212; sum-of-contrigs estimation
Quality coverage: 8.74 in Q20 bases; pulse field gel estimation
Quality coverage: 8.52 in Q20 bases; sum-of-contrigs estimation.

* NOTE: This is a "working draft" sequence. It currently consists of 9 contrigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contrigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1151: contig of 1151 bp in length
* 1152 1251: gap of unknown length
* 1252 1275: contig of 1024 bp in length
* 2276 2375: gap of unknown length
* 2376 2498: contig of 2123 bp in length
* 4499 4599: gap of unknown length
* 4599 6353: contig of 1755 bp in length
* 6354 6453: gap of unknown length
* 6454 8079: contig of 1626 bp in length
* 8080 8179: gap of unknown length
* 8180 30281: contig of 22102 bp in length
* 30282 30381: gap of unknown length
* 30382 65925: contig of 35344 bp in length
* 65926 66025: gap of unknown length
* 66026 107670: contig of 41645 bp in length
* 107671 107771: gap of unknown length
* 107772 167012: contig of 59242 bp in length.

```

FEATURES

source
1. 167012
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CITC-308K20"
/clone_1lb="caltech human BAC library C"

BASE COUNT 43272 a 39820 c 39446 g 43467 t 1007 others
ORIGIN

Query Match 88.1%; Score 1123.8; DB 2; Length 167012;
Best Local Similarity 95.3%; Pred. No. 1.1e-249;
Matches 1192; Conservative 0; Mismatches 52; Indels 7; Gaps 3;

```

QY 24 TGTGAGTCGGCGCGCGGCTGCTGTGCTGTGCGACGGCGCGGAGACTCGC 83
Db 74291 TGTGAGTCGGCGCGCGGCTGCTGTGCTGTGCGACGGCGCGGAGACTCGC 74350
QY 84 GCACATGAGGGGCTTCAGACGCGAGGAGCGCGCGGCTTCCTCCGTGAGACGAG 143
Db 74351 GCACATGAGGGGCTTCAGACGCGAGGAGCGCGCGGCTTCCTCCGTGAGACGAG 74410
QY 144 TCTTCCTCAAAAATGAGAAAGACATATATATCTCCATTTCAATGATTTCAATTATG 203
Db 74411 TCTTCCTCAAAAATGAGAAAGACATATATATCTCCATTTCAATGATTTCAATTATG 74470
QY 204 CAGATTAAGAGTGTTCACATGCTAGTGAATGACACGCTGCTTAATGCAAAAATG 263
Db 74471 CAGATTAAGAGTGTTCACACCTGTGATGGAATGACACCTGTGCTTAATGCAAAAATG 74530
QY 264 AGATTGCTACAAAGACCCCTTTAAACCTATTAAACAAGATGTGAAAAAGAAACACTC 323

```

```

Db 74531 AGTTGCTACAAAGACCCCTTTAAACCCATTAAACCAAGATGTGAAAAAGAAACACTC 74590
QY 324 GGTATGTTGCGAATTTGTTCCGATTAAGATATATCTGGAACACTATGTCGCATCCCTC 383
Db 74591 GGTATGTTGCGAATTTGTTCCGATTAAGATATATCTGGAACACTATGTCGCATCCCTC 74650
QY 384 AGACTTGGAGAACCCAGGCGACAAATGATTAACATATCTGCGCTGTGTGATACATGACC 443
Db 74651 AGACTTGGAGAACCCAGGCGACAAATGATTAACATATCTGCGCTGTGTGATACATGACC 74710
QY 444 CAATTGATGTCTGTAAATTTGGAACGATATGTCGAAGAGTGAATTAATTTGCGCTGA 503
Db 74711 CAATTGATGTCTGTAAATTTGGAACGATATGTCGAAGAGTGAATTAATTTGCGCTGA 74770
QY 504 AAGTTCTAGGCATATTTGGCTATGATTTGACGAAGGGGAAACCGACTGGAAGTATGCGCA 563
Db 74771 AAGTTCTAGGCATATTTGGCTATGATTTGACGAAGGGGAAACCGACTGGAAGTATGCGCA 74830
QY 564 TTAATGTGATGATCCTGATGACGCCAATTAATGATATCAATGATGTCGAAGCGCTGA 623
Db 74831 TTAATGTGATGATCCTGATGACGCCAATTAATGATATCAATGATGTCGAAGCGCTGA 74890
QY 624 AACCTGGCTACTTAAGACCTAGCTGAGTGTGATTAAGAGTATAGAGTTCCTGATGGA 683
Db 74891 AACCTGGCTACTTAAGACCTAGCTGAGTGTGATTAAGAGTATAGAGTTCCTGATGGA 74950
QY 684 AACCAAGAAATGATTTGCTTAATGACGAATTTAAAGATTAAGACCTTGGCAATTTGATA 743
Db 74951 AACCAAGAAATGATTTGCTTAAATGACGAATTTAAAGATTAAGACCTTGGCAATTTGATA 75009
QY 744 TTATTTAAAGCACTATGACACATTTGAAGCAATTTGATGATTAAGAAAGATTAAGAAAG 803
Db 75010 TTATTTAAAGCACTATGACACATTTGAAGCAATTTGATGATTAAGAAAGATTAAGAAAG 75069
QY 804 GAATCAGTTGATGATTAACACTTTTGTGAGAGCCCTTCAAGTGTGATTCGTATGCTG 863
Db 75070 GAATCAGTTGATGATTAACACTTTTGTGAGAGCCCTTCAAGTGTGATTCGTATGCTG 75129
QY 864 CCAGAGCCATGTGATGCTTTTACACACACCTGTAATCTGCTCCACAGATACCAAGC 923
Db 75130 CCAGAGCCATGTGATGCTTTTACACACACCTGTAATCTGCTCCACAGATACCAAGC 75189
QY 924 ACCTGATTAAGTGTTCATCAACCAAGAAATTAAGAGATTTCTGTGAATTCAGACTG 983
Db 75190 ACCTGATTAAGTGTTCATCAACCAAGAAATTAAGAGATTTCTGTGAATTCAGACTG 75249
QY 984 ATATTGCTACATGCTGTTTCACTGATGATTAATGAAGTAAAGTACTAGCTTTCAAGC 1043
Db 75250 ATATTGCTACATGCTGTTTCACTGATGATTAATGAAGTAAAGTACTAGCTTTCAAGC 75308
QY 1044 TTTAATTTGTAGACATCTAATCAATTAAGTAAATTTCTGCTGACTAATCCAAATPACT 1103
Db 75309 TTTAATTTGTAGACATCTAATCAATTAAGTAAATTTCTGCTGACTAATCCAAATPACT 75363
QY 1104 CAGAAATGTTATTCATCTTAAGCAATTTTTCATATCTCAATCAATCAATCAATTTAGCAATG 1163
Db 75364 CAGAAATGTTATTCATCTTAAGCAATTTTTCATATCTCAATCAATCAATCAATTTAGCAATG 75423
QY 1164 CTTAATATATCAAGACGTTGCTATTTGGAAGTCACTTGTAAATATGATGTCGAAGGAGC 1223
Db 75424 CTTAATATATCAAGACGTTGCTATTTGGAAGTCACTTGTAAATATGATGTCGAAGGAGC 75483
QY 1224 ACATATTGAGATGATATGTTATACCATATGTTAGGAATAATATATTTTGTCT 1274
Db 75484 ACATATTGAGATGATATGTTATACCATATGTTAGGAATAATATATTTTGTCT 75534

```

RESULT 11

AC021159 AC021159 213025 bp DNA HTG 19-JUN-2000
LOCUS Homo sapiens chromosome 5 clone RP11-536N17, WORKING DRAFT
DEFINITION SEQUENCE, 41 unordered pieces.

ACCESSION AC021159 GI:8572521
 VERSION AC021159.4 HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 213025)
 AUTHORS Waterston,R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 213025)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 COMMENT On Jun 19, 2000 this sequence version replaced gi:8570288.

----- Genome Center -----
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 41 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1384: contig of 1384 bp in length
 1385 1484: gap of unknown length
 1485 3382: contig of 1898 bp in length
 3383 3482: gap of unknown length
 3483 5013: contig of 1531 bp in length
 5014 5113: gap of unknown length
 5114 6322: contig of 1209 bp in length
 6323 6422: gap of unknown length
 6423 7736: contig of 1314 bp in length
 7737 7836: gap of unknown length
 7837 9198: contig of 1362 bp in length
 9199 9298: gap of unknown length
 9299 11720: contig of 2422 bp in length
 11721 11820: gap of unknown length
 11821 13496: contig of 1678 bp in length
 13497 13598: gap of unknown length
 13599 15122: contig of 1524 bp in length
 15123 15223: gap of unknown length
 15224 17333: contig of 2110 bp in length
 17334 17432: gap of unknown length
 17433 19823: contig of 2391 bp in length
 19824 19923: gap of unknown length
 19924 21811: contig of 1888 bp in length
 21812 21911: gap of unknown length
 21912 24564: contig of 2653 bp in length
 24565 24664: gap of unknown length
 24665 27708: contig of 3044 bp in length
 27709 27808: gap of unknown length
 27809 30176: contig of 2368 bp in length
 30177 30276: gap of unknown length
 30277 33515: contig of 3239 bp in length
 33516 33615: gap of unknown length
 33616 37205: contig of 3590 bp in length
 37206 37305: gap of unknown length
 37306 39846: contig of 2541 bp in length
 39847 39946: gap of unknown length
 39947 42384: contig of 2438 bp in length
 42385 42484: gap of unknown length
 42485 46260: contig of 3776 bp in length
 46261 46360: gap of unknown length
 46361 49107: contig of 2747 bp in length
 49108 49207: gap of unknown length
 49208 54501: contig of 5294 bp in length
 54502 54601: gap of unknown length
 54602 59315: contig of 4714 bp in length

59316 59415: gap of unknown length
 59416 65627: contig of 6212 bp in length
 65628 65727: gap of unknown length
 65728 72486: contig of 6758 bp in length
 72487 72585: gap of unknown length
 72586 77910: contig of 5325 bp in length
 77911 78010: gap of unknown length
 78011 84348: contig of 6338 bp in length
 84349 84448: gap of unknown length
 84449 90437: contig of 5989 bp in length
 90438 90537: gap of unknown length
 90538 97928: contig of 7391 bp in length
 97929 98028: gap of unknown length
 98029 105065: contig of 7037 bp in length
 105066 105165: gap of unknown length
 105166 111323: contig of 6158 bp in length
 111324 111423: gap of unknown length
 111424 117735: contig of 6311 bp in length
 117736 117834: gap of unknown length
 117835 125427: contig of 7593 bp in length
 125428 125527: gap of unknown length
 125528 134991: contig of 9464 bp in length
 134992 135091: gap of unknown length
 135092 143447: contig of 8356 bp in length
 143448 143547: gap of unknown length
 143548 156094: contig of 12547 bp in length
 156095 156194: gap of unknown length
 156195 157939: contig of 1745 bp in length
 157940 158039: gap of unknown length
 158040 175011: contig of 16972 bp in length
 175012 175111: gap of unknown length
 175112 187724: contig of 12613 bp in length
 187725 208210: contig of 20366 bp in length
 208211 208311: gap of unknown length
 208312 213025: contig of 4715 bp in length.

FEATURES

source

misc_feature
 1 1384
 /note="assembly_name:Contig11"
 misc_feature
 1485..3382
 /note="assembly_name:Contig14"
 misc_feature
 3483..5013
 /note="assembly_name:Contig15"
 misc_feature
 5114..6322
 /note="assembly_name:Contig16"
 misc_feature
 6423..7736
 /note="assembly_name:Contig18"
 misc_feature
 7837..9198
 /note="assembly_name:Contig24"
 misc_feature
 9299..11720
 /note="assembly_name:Contig25"
 misc_feature
 11821..13498
 /note="assembly_name:Contig26"
 misc_feature
 13599..15122
 /note="assembly_name:Contig27"
 misc_feature
 15223..17332
 /note="assembly_name:Contig28"
 misc_feature
 17433..19823
 /note="assembly_name:Contig30"
 misc_feature
 19924..21811
 /note="assembly_name:Contig31"
 misc_feature
 21912..24564
 /note="assembly_name:Contig32"
 misc_feature
 24665..27708
 /note="assembly_name:Contig33"
 misc_feature
 27809..30176
 /note="assembly_name:Contig34"
 misc_feature
 30277..59315

```
/note="assembly_name:Contig36"
33616..37205
/note="assembly_name:Contig37"
37306..39846
/note="assembly_name:Contig38"
39947..42384
/note="assembly_name:Contig39"
42485..46260
/note="assembly_name:Contig40"
46361..49107
/note="assembly_name:Contig41
clone_end:SP6
vector_side:left"
49208..54501
/note="assembly_name:Contig42"
54602..59315
/note="assembly_name:Contig43"
59416..65627
/note="assembly_name:Contig44"
65728..72485
/note="assembly_name:Contig45"
72586..77910
/note="assembly_name:Contig46"
78011..84348
/note="assembly_name:Contig47"
84449..90437
/note="assembly_name:Contig48"
90538..97928
/note="assembly_name:Contig49"
98029..105065
/note="assembly_name:Contig50"
105166..111323
/note="assembly_name:Contig51"
111424..117734
/note="assembly_name:Contig53"
117835..125427
/note="assembly_name:Contig54"
125528..134991
/note="assembly_name:Contig55"
135092..143447
/note="assembly_name:Contig56"
143548..156094
/note="assembly_name:Contig57"
156195..157939
/note="assembly_name:Contig58"
158040..175011
/note="assembly_name:Contig59"
175112..187724
/note="assembly_name:Contig60"
187825..208210
/note="assembly_name:Contig61"

Query Match      88.1% Score 1123.8; DB 2; Length 213025;
Best Local Similarity 95.3%; Pred. No. 1.2e-249;
Matches 1192; Conservative 0; Mismatches 52; Indels 7; Gaps 3;

misc_feature      /note="assembly_name:Contig36"
misc_feature      /note="assembly_name:Contig37"
misc_feature      /note="assembly_name:Contig38"
misc_feature      /note="assembly_name:Contig39"
misc_feature      /note="assembly_name:Contig40"
misc_feature      /note="assembly_name:Contig41
misc_feature      /note="assembly_name:Contig42"
misc_feature      /note="assembly_name:Contig43"
misc_feature      /note="assembly_name:Contig44"
misc_feature      /note="assembly_name:Contig45"
misc_feature      /note="assembly_name:Contig46"
misc_feature      /note="assembly_name:Contig47"
misc_feature      /note="assembly_name:Contig48"
misc_feature      /note="assembly_name:Contig49"
misc_feature      /note="assembly_name:Contig50"
misc_feature      /note="assembly_name:Contig51"
misc_feature      /note="assembly_name:Contig53"
misc_feature      /note="assembly_name:Contig54"
misc_feature      /note="assembly_name:Contig55"
misc_feature      /note="assembly_name:Contig56"
misc_feature      /note="assembly_name:Contig57"
misc_feature      /note="assembly_name:Contig58"
misc_feature      /note="assembly_name:Contig59"
misc_feature      /note="assembly_name:Contig60"
misc_feature      /note="assembly_name:Contig61"

Query 24 TGTCACTGGGCGCGGCTGCTGCTGTCAGCGCGCGGCGAGACTCG 83
          |||
Db 75573 TGTCACTGGGCGCGGCTGCTGCTGTCAGCGCGGCGAGACTCG 75632

Query 84 GCACATGAGCGGCTGACACGAGAGCGCGCGGCTTCTCCCTGAGTACCGAG 143
          |||
Db 75633 GCACATGAGCGGCTGACACGAGAGCGCGCGGCTTCTCCCTGAGTACCGAG 75692

Query 144 TCTTCTCAAAAATGAGAAAGCATATATATCTCCATTTTCATGATTCATTTTATG 203
          |||
Db 75693 TCTTCTCAAAAATGAGAAAGCATATATATCTCCATTTTCATGATTCATTTTATG 75752

Query 204 CAGATAGAGATGTTTCAACATGTTAGTGAAGTACACGCTGCTTAATGCAAAATG 263
          |||
Db 75753 CAGATAGAGATGTTTCAACATGTTAGTGAAGTACACGCTGCTTAATGCAAAATG 75812

Query 264 AGATGCTCAAAAGACCCCTTTAAACCTATTATTAACAAGATGGAAGAAAGAACTTC 323
```

```
Db 75813 AGTTCCTCAAAAGACCCCTTTAAACCCCTTTAAACAGATGTGAAAAAAGAACTTC 75872
          |||
Query 324 GGTATGTCGCAATTTGTTCCCGCTATPAAAGATATATCTGAGAACTATGTCATCCCTC 383
          |||
Db 75873 GGTATGTCGCAATTTGTTCCCGCTATPAAAGATATATCTGAGAACTATGTCATCCCTC 75932

Query 384 AGACTTGGAGAACCCAGGCGACAAATGATPAAACATACCTGCTGTTGCTGCAATGACC 443
          |||
Db 75933 AGACTTGGAGAACCCAGGCGACAAATGATPAAACATACCTGCTGTTGCTGCAATGACC 75992

Query 444 CAATTGATGTGTGTAATTTGAGACAGATATGTCAGAGAGTAAATTTGGCGTGA 503
          |||
Db 75993 CAATTGATGTGTGTAATTTGAGACAGATATGTCAGAGAGTAAATTTGGCGTGA 76052

Query 504 AAGTCTAGCATATTTGCTATGATTTGAGAGAGGGAACCCAGTGAAGCATTTGCCA 563
          |||
Db 76053 AAGTCTAGCATATTTGCTATGATTTGAGAGAGGGAACCCAGTGAAGCATTTGCCA 76112

Query 564 TTAATGTGATGATCCTGATGCGACCAATTAATGATATCAATGATGCAACGCGCTGA 623
          |||
Db 76113 TTAATGTGATGATCCTGATGCGACCAATTAATGATATCAATGATGCAACGCGCTGA 76172

Query 624 AACCTGCTACTTGAAGCTACTGTGAGCTGTTTGAAGTATPAAAGTTCTGATGGA 683
          |||
Db 76173 AACCTGCTACTTGAAGCTACTGTGAGCTGTTTGAAGTATPAAAGTTCTGATGGA 76232

Query 684 AACCGAAATGAGTTTGGCTTTAATGCGAATTTAAAGATTAAGGACTTGCATTTGATA 743
          |||
Db 76233 AACCGAAATGAGTTTGGCTTTAATGCGAATTTAAAGATTAAGGACTTGCATTTGATA 76291

Query 744 TTTATTAAGACATCATGACCATTTGAAAGCATTTAGTGAATGAAGAAAGTGAAG 803
          |||
Db 76292 TTTATTAAGACATCATGACCATTTGAAAGCATTTAGTGAATGAAGAAAGTGAAG 76351

Query 804 GAATCAGTTGCAATGAAATACAACTTTGTCTGAGAGCCCTTCAAGTGTATCTGATGCTG 863
          |||
Db 76352 GAATCAGTTGCAATGAAATACAACTTTGTCTGAGAGCCCTTCAAGTGTATCTGATGCTG 76411

Query 864 CCAGAGCCATTTGTGATGCTTTTACCAACCACTGTAATCTCCCTGACAGTACCAAGC 923
          |||
Db 76412 CCAGAGCCATTTGTGATGCTTTTACCAACCACTGTAATCTCCCTGACAGTACCAAGC 76471

Query 924 AGCTGATATAGTGTCTTCATCACCAGAAAACTAATGATTTCTTGAATATCAAGCTG 983
          |||
Db 76472 AGCTGATATAGTGTCTTCATCACCAGAAAACTAATGATTTCTTGAATATCAAGCTG 76531

Query 984 ATATTGCTCATCGTGTTCATCTGATGATTTAGAAATGAAGTATGATCTTTTCAAGC 1043
          |||
Db 76532 ATATTGCTCATCGTGTTCATCTGATGATTTAGAAATGAAGTATGATCTTTTCAAGC 76590

Query 1044 TTTAAATTTGTAGAACTCATTAACATAAGTAATTTCTGCTGACTAATTCATTAAT 1103
          |||
Db 76591 TTTAAATTTGTAGAACTCATTAACATAAGTAATTTCTGCTGACTAATTCATTAAT 76645

Query 1104 CAGATGTTATCCATCTAAGATATTTTTCATATTCACATAAGTATTAATCTTAAAGC 1163
          |||
Db 76646 CAGATGTTATCCATCTAAGATATTTTTCATATTCACATAAGTATTAATCTTAAAGC 76705

Query 1164 CTTAATATCAAAAGCAGTTGATTTGGAAGTCACTTGAATAGATGCAAGGGAGC 1223
          |||
Db 76706 CTTAATATCAAAAGCAGTTGATTTGGAAGTCACTTGAATAGATGCAAGGGAGC 76765

Query 1224 ACATATTTGATATATATTTTACCATATATTTAGGAAATTAATTTTGTCT 1274
          |||
Db 76766 ACATATTTGATATATATTTTACCATATATTTAGGAAATTAATTTTGTCT 76816

RESULT 12
AC011012 155645 bp DNA HTG 03-NOV-2000
LOCUS AC011012 155645 bp DNA HTG 03-NOV-2000
DEFINITION Homo sapiens clone RP11-8D23, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
```


Db 1022 TTTAAATTATAGACTAATC-----AGTCAGATTCTGCTGACCAATCCAGTGTATT 1075

Qy 1104 CAGATGTATCCATCTAAGCATTTTTCATATCTCAACTAAGATTAAGTTTACACATG 1163

Db 1076 CATCATTTATCTATCTAAGACATTATC--GTGCACTAAGGATGACATTTAGTTCTG 1133

Qy 1164 CTTAAATATCAAGAGCTTGTATTTGAGAGTCTGTGAATAGATGCAAGGGAGC 1223

Db 1134 CTTAAATATCAAGGGCTTGTAGTTGTTGAGAGTCTGTGAAGATGCAAGATAGT 1193

Qy 1224 ACATTTGATGTATATCTTACCATATCTTGAAGATTAATTTATTT 1271

Db 1194 ACATTTGATGTATATCTTGTCTGATGTTAGGAATTAATTTATTT 1241

RESULT 14

LOCUS AF108211 846 bp mRNA PRI 14-Apr-1999

DEFINITION Homo sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds.

ACCESSION AF108211

VERSION AF108211.1 GI:4583152

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 846)

AUTHORS Rumsfeld,J., Ziegelerbauer,K. and Spaltmann,F.

TITLE Cloning, expression, affinity purification and characterization of polyhistidine-tagged cytosolic Saccharomyces cerevisiae and human inorganic pyrophosphatases for differential screening of compounds for antifungal activity

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 846)

AUTHORS Rumsfeld,J., Ziegelerbauer,K. and Spaltmann,F.

TITLE Direct Submission

JOURNAL Submitted (20-NOV-1998) Research Antinfectives 1, Bayer AG, P.O. Box 10179, Wuppertal 42096, Germany

FEATURES

location/Qualifiers

1. 846

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_line="BjAB"

/cell_type="B-cell"

<1..>846

/note="iPpase"

/codon_start=1

/product="cytosolic inorganic pyrophosphatase"

/protein_id="AAD24964.1"

/db_xref="GI:4583153"

/translation="STERRAASFLEVRPLKNGKGVISPEHDIPIYADKDFHNV EYPRNSAKMEIATKDPINPIKQDYKKKRLVNLPPYKGIINNYGATPOTWEDPGH NDKHTGCCGDMPIIDVCEIGSKVARGELIGKVLGIAMIDESETDKVIAINVDPA DAANYNDINDVRLKPGYLEAVTVMFRYKVPDGRKPEMEFNAEFKDKDAIDILKS THDKMALVTKTKNGKISCNMTLSESPKCDPDAAARAIVDALPPCESACTVPTDV DKWEHH"

CDS

BASE COUNT 270 a 155 c 201 g 220 t

ORIGIN

Query Match 66.0%; Score 841.2; DB 9; Length 846;

Best Local Similarity 99.6%; Pred. No. 2e-184;

Matches 843; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 101 AGCACCAGAGAGCGCGCGCCCTTCTCCCTGAGTACCGAGTCTTCTCAAAAATGAG 160

Db 1 AGCACCAGAGAGCGCGCGCCCTTCTCCCTGAGTACCGAGTCTTCTCAAAAATGAG 60

Qy 161 AAAGACATATATATCTCATTTATGATATTTCAATTTATGAGATAGAGATGTTT 220

Db 61 AAAGACATATATATCTCATTTATGATATTTCAATTTATGAGATAGAGATGTTT 120

Qy 221 CACATGTAGTGAAGTACCAAGCTGTGCTATATGCAAAAATGAGATTGCTCAAAAGAC 280

Db 121 CACATGTAGTGAAGTACCAAGCTGTGCTATATGCAAAAATGAGATTGCTCAAAAGAC 180

Qy 281 CTTTAAACCTTATTAACAGATGTGAAAAAGAAAACTTGCATGTTCCGAATTTG 340

Db 181 CTTTAAACCTTATTAACAGATGTGAAAAAGAAAACTTGCATGTTCCGAATTTG 240

Qy 341 TTCCCGTATTAAGATATATCTGTAAGTATGTTGTCATCCCTCAGACTGGGAAGCCCA 400

Db 241 TTCCCGTATTAAGATATATCTGTAAGTATGTTGTCATCCCTCAGACTGGGAAGCCCA 300

Qy 401 GGGCACAATGATTAACATACCTGCTGTGTGTGTGCAATGACCAATGATGTGTGAA 460

Db 301 GGGCACAATGATTAACATACCTGCTGTGTGTGTGCAATGACCAATGATGTGTGAA 360

Qy 461 ATTGAAGCAAGTATGTGCAAGAGTGAATTAATTTGGCTAAAGTTCTAGGCATTTG 520

Db 361 ATTGAAGCAAGTATGTGCAAGAGTGAATTAATTTGGCTAAAGTTCTAGGCATTTG 420

Qy 521 GCTATGATTGAGCAAGGGGAACCCGACCTGGAAGTATTCCTTAATGTGATGATCT 580

Db 421 GCTATGATTGAGCAAGGGGAACCCGACCTGGAAGTATTCCTTAATGTGATGATCT 480

Qy 581 GATGAGCCCAATTATTAATGATATCAATGATGTCAACCGCTGAACCTGCTACTTAGAA 640

Db 481 GATGAGCCCAATTATTAATGATATCAATGATGTCAACCGCTGAACCTGCTACTTAGAA 540

Qy 641 GCTACTGTGACTGCTTTAAGAGTATTAAGTCTTCTGATGAGAAACCAAGAAATGACTTT 700

Db 541 GCTACTGTGACTGCTTTAAGAGTATTAAGTCTTCTGATGAGAAACCAAGAAATGACTTT 600

Qy 701 GCGTTAATGCAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTAT 760

Db 601 GCGTTAATGCAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTAT 660

Qy 761 GACCATTTGGAAGCAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTAT 820

Db 661 GACCATTTGGAAGCAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTAT 720

Qy 821 ACACTTTGTCTGAGAGCCCTTCAAGTGTATCTGATGCTGCCAGACCATTTGGCAT 880

Db 721 ACACTTTGTCTGAGAGCCCTTCAAGTGTATCTGATGCTGCCAGACCATTTGGCAT 780

Qy 881 GCTTACCAACCACTGTGATCTGCTGACAGTACCAACAGAGCTGATAGTGTTC 940

Db 781 GCTTACCAACCACTGTGATCTGCTGACAGTACCAACAGAGCTGATAGTGTTC 840

Qy 941 CATCAC 946

Db 841 CATCAC 846

RESULT 15

LOCUS BC010468 1243 bp mRNA ROD 12-JUL-2001

DEFINITION Mus musculus, RIKEN cDNA 2010317E03 gene, clone MGC:6716

ACCESSION BC010468

VERSION BC010468.1 GI:14714656

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1243)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

COMMENT Contact: MGC help desk

Email: cgarabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E Consortium (ILLUMI)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.tmc.edu/cdna/>
Contact: villalobebcm.tmc.edu
Villalob, D.K., Luna, R.A., Hale, S.M., Huiyk, S., Lu, X., Garcia,
A.M., Holloman, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Mozny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Series: 10 Row: F Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: Genomscan gene prediction, Similarity but not identity to protein, *StatComp* filters

FEATURES
SOURCE

```

/organism="Mus_musculus"
/db.xref="locusID:67895"
/db.xref="taxon:10090"
/clonename="MCC:5716 IMAGE:3565780"
/tissue_type="Mammary tumor. Metallotionien-TGF alpha
model: 10 month old virgin mouse. Taken by biopsy."
/clonename="NCI-CGAP_Maml1"
/lab host="DH10B"
/note="Vector: pCMV-Sport6"
42. .911

```

CDS

```

/codon_start=1
/product="Riken CDNA 2010317E03 gene"
/protein_id="AAH10468.1"
/db_xref="GI:14714657"
/translation="MSGSSSERAAPFLETVRFVFLKNGKGVISPEHDVPIYADKDFV
HAAVEAFKRWSSNAKMEIATKDDLPNLPKQVKKKGLRYANLIPFYGIWNVGAILPOTWE
DPGHSDHTTGCAGDDPLDVCIEIGSKVARSSEILVKKVLGLLAMDSEETMKKATAI
VDDPAAANYKDISDVERLPGVLEATVYMFRRYKYPDPKPNERRAFNNEKNKDPFAVI
IKSHDYKALKALVTKRTKDGKISCMNTVSESPEFCDDDAKAIADLPPCESACSSCI
PLDVDPKMFHCKKN"

```

QY	410	GATTAACATCTGGCTGTGTGTGACAAATGACCCCAATGTATGTGTGTGAATTTGAAG	469
Db	363	GACAAAGCACTGGCTGTGTGTGTGACAAAGCCCAATCATGATTTTGTGAATCGGAAG	422
QY	470	AAGTATGTSCAAGAGGTAAATTAATGTGGCGGAAGAGTTTACGCATATTGGCTATGATT	529
Db	423	AAGTGTGTCCAGAGGTAGATTAATTCAGGGTGAAGGTCGTGGCCATACGTGGCCATATGC	482
QY	530	GACGAAGGGAAACCGACTGGAAAGTCATTGGCATTAAATGTGATGATCCGTGATGACGCC	589
Db	483	GATGAGGGCCAGACCCACTGGAAAGTCATTGGCATTAAATGTGAGCATCCGGACCGACCC	542
QY	590	AATATATATGATPACATGATGTCAACAGCGCTGAACACCTGGCTACTTGAAGACTACTGTG	649
Db	543	AATATTAAGATTAATTTCTGATGTCAAGCGCTTAACACCGCGTATATTGAAGTACTGTT	602
QY	650	GACGTGTTTAGAAGTATAAGTTCCTGATGTGAAAAACCAAAATGATTCGCTTTAAT	709
Db	603	GATTGTGTTTAGAAGGTATAAGTTCCTGATGTGAAAAACCAAAATGATTCGCTTTAAT	662
QY	710	GCAGATTTTAAGATTAAGGACTTGGCATTTGATATTAATAAAGCAGCATGACCATTTGG	769
Db	663	GCAGATTTTAATAAACAAGACTTTGCAAGTTGACATTATTAATAAGCACCCAGATTAATGG	722
QY	770	AAAGCATTAAGTACTTAAGAAAAAGCAATGAAAAAGATCAGTTGCATGATCAACTTTG	829
Db	723	AAAGCATTAAGTACTTAAGAAAAAGCAATGAAAAAGATCAGTTGCATGATCAACTTTG	782
QY	830	TCTGAGACCCCTTCAGTGTGATGATCCGATGATGCTGCAGAGCCCATTTGTGATCTTTACA	889
Db	783	TCTGAGACCCCTTCAGTGTGATGATCCGATGATGCTGCAGAGCCCATTTGTGATCTTTGCA	842
QY	890	CCACCCCTGTGATCTGCTGCACAGTACCAACAGACGCTGATTAAGTGGTTCATACACAG	949
Db	843	CCACCATGTGATCTGCTGCTGCACAGTACCAACAGATGATGATTAAGTGGTTCATACAGG	902
QY	950	AAAAACTATGATATTTCTCTGGAATACAGAGTGATTAATGCTATCATCGTTCATCTGGA	1009
Db	903	AAAAACTGAAAGGCTCTCTCTGGAACAGAGCTGTTAAGGCCGCAATCCGTTCACCTGGA	962
QY	1010	TGATATTAGAGTAAGATAGTAGTACCTTTTCAAGAGCTTTAAATTTGTAGAATCATCTAACT	1065
Db	963	AGCATTAAGTGAAGAAAGA-----TTTTCAGAGCTTAATTTCTGTACTCAGCGAA-T	1015
QY	1070	AAAGTAATTTCTGTGTGACTAATATCCAAATATATCTAGAAATGATATCCATCTAAAGATT	1122
Db	1016	GAAGTAGACTGTGTGTGACTGTGCTCAGCTACTATTCAGAACTGGTTCAGAAAGTTAGCATC	1075
QY	1130	T-----TCATATCTCAACTAAGTATACCTTTTAGCAGCATGCTTAAATATCAAAAGCATT	1183
Db	1076	TAGACACCTCAATCTCAACCAAGACAGCTTTTATTA- AACCTTAGCTCCAAAGCGCTG	1133
QY	1183	GTCATTTTGGAAGTCACTTGTGTGATAGATGTCCAAAGGGAGACATATTTGGATGTATATGT	1241
Db	1135	GGGACATGGAACATCTTTGTGTGATGAGATGTCACTGTGAGACACCATTTGATGTATCATGT	1193
QY	1243	TACCATATGTTAGAAATTAATTTTGTGGTG 1275	
Db	1195	TTCTGTGTGAGAAATTAATTTTTTTACTG 1226	

Search completed: March 29, 2002, 19:30:55
Job time: 2298 sec

Query Match	99.98;	Score 1274;	DB 2;	Length 1275;
Best Local Similarity	100.08;	Pred. NO. 0;		
Matches 1275;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

Db	1141	ACTAAGTAACCTTTTATGACATGCTTAATAATATCAAGCAGGTTGTCATTTGGAAATCAGT	1200
Qy	1201	GTGAATGATGTCGAACGGGAGACAAATTTGATATATTTTACCATTATGTTAGCAAT	1260
Db	1201	GTGAATGATGTCGAACGGGAGACAAATTTGATATATTTTACCATTATGTTAGCAAT	1260
Qy	1261	AAAATTATTTTCTGCTG	1275
Db	1261	AAAATTATTTTCTGCTG	1275

Patent No. 5670367

APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.

```

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner

ADDRESS: Foley & Larimer
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria

CITY: Alexandria
STATE: VA

COUNTRY: USA
ZIP: 22313-0299

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

```

; MEDIUM TYPE: floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463

```

IN RE: **CLASSIFICATION: 435**
 FILING DATE: 02/06/2025
 CLASSIFICATION: 435

CLASSIFICATION: 435
PRIOR APPLICATION DATA: 435

APPLICATION NUMBER: US/07/935,313
FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991

FILED DATE: 20-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT C+eeba ?

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-93
TELEFAX: (703) 683-4109

TELEFAX: (703) 003 4105
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:

; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 7318 base pairs
 ;

```

; LENGTH: 7218 base pairs
; TYPE: 'nucleic acid'

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
;
```

IMMEDIATE SOURCE:
CLONE: RT2a9t-F1a

CLONE: pzqpt-f1s
US-08-232-463-14

Query Match 4.08;
Best Local Similarity 3.78; P

Matches 15; Conservative 223

429 GTGGACATGACCAATTGATG

Db 1463 GTAGTTAAGAGATAGAGAATT

489 AAATAATTGCCGTGAAGTCTAG

07	409	AAAAAA110000100000110100
0b	1403	BBBBBBBBBBBBBBBBBBBBBBBBBB

D6 **1403 RRRRRRRRRRRRRRRRRRRRRR**

RESULT 4
US-08-809-267-9
; Sequence 9, Application US/08809267
; Patent No. 5861296

Query Match	3.1%;	Score 39.2;	DB 2;	Length 531;
Best Local Similarity	50.5%;	Pred. No. 0.1;		

RESULT 5
PCT-US95-13662A-9
; Sequence 9, Application PC/TUS9513662A
; GENERAL INFORMATION.

Query Match	3.1%	Score 39.2;	DB 5;	Length 531;
Best Local Similarity	50.5%;	Pred. No. 0.1;		
Matches 95;	Conservative 0;	Mismatches 93;	Indels 0;	Gaps 0;

OY 425 TGTGTTGGTCAACAATGACCCCATTTGATGTTGTGAATAATTGGAAGCAAGTATGTGCAGA 484
| | | | | | | | | | | | | |
Db 187 TTGTACGATCATGACGACCCTTTTGACATCATGTCATATGAGGAGGCCAACATATCCG 246

Query Match	2.9%	Score 37.6;	DB 3;	Length 1155;
Best Local Similarity	47.8%;	Pred. No. 0.44;		
Matches 109;	Conservative 0;	Mismatches 119;	Indels 0;	Gaps 0

```

Query Match      2.98; Score 36.8; DB 2; length 19124;
Best Local Similarity 46.28; Pred. No. 3.1; 142; Indels 0; Gaps 0;
Matches 122; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
0Y 1012 TATTGAGAGTAAGTAGAGCTTTCAAGCTTTAAATTGTAGAACCACTAAGCTAA 1071
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
TITLE OF INVENTION: WERNER'S SYNDROME
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 207:
SEQUENCE CHARACTERISTICS:
LENGTH: 29604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-207

Query Match: 2.7%; Score 34.8; DB 3; Length 29604;
Best Local Similarity 51.5%; Pred. No. 14;
Matches 105; Conservative 0; Mismatches 97; Indels 2; Gaps 1;
QY 170 TATATATCCATTTTCATGATTCATTAATTCATTAAGCATGCTTTCATCATGTA 229
DB 18026 TCTAAGTGGCAATTAACACATATTAAGAAATGATTAATTCATTAATTCATGACATGTA 17967
QY 230 GTTGAAGTACCAGCTGCTTAATGCA--AAATGAGATTGCTACAAAGACCCCTTAA 287
DB 17966 TTTAACTAGACGGGGTGGAGAAATTAATTAATAAGGATATCATGAGCTTTTAT 17907
QY 288 ACCCTATTAACAAGATGTGAAAAAAGAAATTCGCTATGTTGCAATTTGTTCCGT 347
DB 17906 TTTCTATTATTTTACATTAATAATTAATAGAAAGAACTACTCTGAAAGTTGCTCTG 17847
QY 348 ATAAGATATATCTGCAACTATG 371
DB 17846 ACCAGTATATCTCATCATG 17823

RESULT 15
US-07-715-751B-2/c
Sequence 2, Application US/07715751B
Patent No. 5391725
GENERAL INFORMATION:
APPLICANT: CORUZZI, GLORIA M
APPLICANT: EDWARDS, JANICE W
APPLICANT: WALKER, ELSBETH L
APPLICANT: BREARS, TIMOTHY B
TITLE OF INVENTION: NOVEL ORGAN-SPECIFIC PLANT PROMOTER
TITLE OF INVENTION: SEQUENCES
NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/715,751B
FILING DATE: 19910613
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 3288-017-999
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741
TELEX: 166141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1939 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-715-751B-2

Query Match: 2.7%; Score 34.6; DB 1; Length 1939;
Best Local Similarity 45.4%; Pred. No. 4;
Matches 124; Conservative 0; Mismatches 149; Indels 0; Gaps 0;
QY 1001 TCATCTGATGTATTAGAGTAAGTAAGTAGTAGCTTTTCAAGCTTTAAATTTGTAAGACT 1060
DB 1637 TCACCTTAATTTTCTTCAAAAGAAAGAAAGATATGATATACCTTTGATTTCTTTTCG 1578
QY 1061 CATCTAATTAAGTAATTTCTGCTGACATCAATCAATATACAGATTTATTCATCT 1120
DB 1577 AATCTAATGACATTAATTTGGCGGTGATGATGATCTGCTTAAATGATTAATGATATT 1518
QY 1121 AAAGCAATTTTCATATCTCAACTAAGATTAATCTTTAGCACATGCTTAAATATCAAGCAG 1180
DB 1517 TTATGAGAGCAATGTACGGAAGATGTTTAAATATTAATTAATTAATGATG 1458
QY 1181 TTGTCATTTGGAATCTACTTGTGATATGATGTGCAAGGGGACATATTTGGATGTATAT 1240
DB 1457 GTTGTAGAGATCTTCACTTAATTAATTAATTTTGTGTAAGAACTAATTAATTAAT 1398
QY 1241 GTTACCAATATGTAAGAAATTAATTAATTTTGC 1273
DB 1397 TTTATATTTTCATTCATAAATTAATTTTATGAC 1365

Search completed: March 29, 2002, 19:53:01
Job time: 354. sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 18:53:17 ; Search time 1498.3 Seconds

(without alignments)
9144.275 Million cell updates/sec

Title: US-09-415-540-2
Perfect score: 1275
Sequence: 1 CAAGAGGTTCGGCTCTCT.....GAATATAATTTTCGTCG 1275

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_estl:*
11: gb_estc:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pla:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	933.6	73.2	978	10	AL525341
2	898.8	70.5	985	10	AL574658
3	836.2	65.6	1293	12	AK008575
4	813	63.8	826	10	AL519324
5	811.6	63.7	932	11	BG676397
6	776.8	60.9	826	10	AL515056
7	760	59.6	996	11	BG674343
8	742.8	58.3	773	11	BG715195
9	742	58.2	771	11	BI258591
10	739.8	58.0	772	11	BG713976
11	732.8	57.5	878	11	BI259308
12	730.2	57.3	1308	12	AK020291

13	721	56.5	915	11	BG574468
14	720.4	56.5	845	11	BG708966
15	719.4	56.4	762	11	BG721093
16	712.6	55.9	804	11	BI260339
17	709	55.6	714	2	BG700549
18	707.2	55.5	754	1	BG699988
19	707.2	55.5	1076	11	BG179769
20	706.8	55.4	836	11	BG715251
21	706.2	55.4	955	11	BG704401
22	705.8	55.4	955	11	BF791373
23	705.4	55.3	805	11	BF967952
24	704.2	55.2	769	11	BG702317
25	703.6	55.2	751	10	AI078153
26	691.6	54.2	775	10	AI956149
27	688.8	54.0	741	10	AL550933
28	687	53.9	828	11	BG705302
29	675.4	53.0	711	11	BG777143
30	672.8	52.8	734	11	BE797649
31	672.2	52.7	998	11	BF966851
32	670.2	52.6	781	11	BG108210
33	667.4	52.3	714	11	BG702955
34	663.4	52.0	929	11	BF793034
35	662.2	51.9	775	11	BG036446
36	660	51.8	693	11	BG703149
37	659.6	51.7	847	11	BI093604
38	658.4	51.6	740	11	BG497509
39	654.8	51.5	843	11	BG106064
40	654.8	51.4	905	11	BG027821
41	653	51.2	997	11	BG255623
42	650.8	51.0	659	10	AI744359
43	648.8	50.9	676	11	BG499929
44	648	50.9	921	11	BG282637
45	645.8	50.7	677	10	AL519323

ALIGNMENTS

RESULT 1
AL525341
LOCUS
DEFINITION AL525341 978 bp mRNA
prime, mRNA sequence.
ACCESSION AL525341
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 978)
L.W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNALS
COMMENT
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 978
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODC011YC02"
/clone_id="LTL.NFL003.NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life

Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (41) 301 610
8371 Email : filiang@lifetech.com URL :
http://fulllength.invitrogen.com
BASE COUNT 316 a 169 c 213 g 267 t 13 others
ORIGIN

Query Match 73.2% Score 933.6; DB 10; Length 978;
Best Local Similarity 97.7%; Pred. No. 8.5e-218;
Matches 956; Conservative 10; Mismatches 11; Indels 2; Gaps 2;

255 CAAAAATGAGATTGCTACAAAGAGACCTTTAAACCTTTAAACAGATGTAAG 314
1 CAAAAATGAGATTGCTACAAAGAGACCTTTAAACCTTTAAACAGATGTAAG 60
315 GAAACTTGGCTATGTTGGCAATTTTCCCGTATAAGGATATCTGAACTATG 374
61 GAAACTTGGCTATGTTGGCAATTTTCCCGTATAAGGATATCTGAACTATG 120
375 CCATCCCTGAGCTTGGGAGAACCCGAGCAGATGATTAACACTGCTGTGGTG 434
121 CCATCCCTGAGCTTGGGAGAACCCGAGCAGATGATTAACACTGCTGTGGTG 180
435 ACAATGACCAATTGATGTGTGTAATTTGAAGCAAGTATGCAAGGTGAATA 494
181 ACAATGACCAATTGATGTGTGTAATTTGAAGCAAGTATGCAAGGTGAATA 240
495 TTGGGGTGAAATTTCTAGCAATTTGGCTATATTGACGAAGGGGAACCCAG 554
241 TTGGGGTGAAATTTCTAGCAATTTGGCTATATTGACGAAGGGGAACCCAG 300
555 TCATTGCCCTTATGATGGATGATCCGATGACGCAATTAATATGATCATGAT 614
301 TCATTGCCCTTATGATGGATGATCCGATGACGCAATTAATATGATCATGAT 360
615 AACGGCTGAAACCTGCTACTTAGAAGCTACTGTGACTGTTTGAAGATTAAG 674
361 AACGGCTGAAACCTGCTACTTAGAAGCTACTGTGACTGTTTGAAGATTAAG 420
675 CTGATGGAAGCAAGCAAAATGAGTTGGCTTTAATGCAAGATTTAAGTAAG 734
421 CTGATGGAAGCAAGCAAAATGAGTTGGCTTTAATGCAAGATTTAAGTAAG 480
735 CCATTGATATTTAATTAAGCACTCATGACCATTTGAAAGCATTAAGTAAAG 794
481 CCATTGATATTTAATTAAGCACTCATGACCATTTGAAAGCATTAAGTAAAG 540
795 ATGAAAAGGAATCAGTTGATGATGATGATGATGATGATGATGATGATGAT 854
541 ATGAAAAGGAATCAGTTGATGATGATGATGATGATGATGATGATGATGAT 600
855 CTGATGCTGCAAGACCTGTTGATGATGATGATGATGATGATGATGATGAT 914
601 CTGATGCTGCAAGACCTGTTGATGATGATGATGATGATGATGATGATGAT 660
915 TACCAAGAGAGTGTAGTGTGATGATGATGATGATGATGATGATGATGAT 974
661 TACCAAGAGAGTGTAGTGTGATGATGATGATGATGATGATGATGATGAT 720
975 TACCAAGAGAGTGTAGTGTGATGATGATGATGATGATGATGATGATGAT 1034
721 TACCAAGAGAGTGTAGTGTGATGATGATGATGATGATGATGATGATGAT 780
1035 TTTCAGGCTTAAATTTGATGATGATGATGATGATGATGATGATGATGAT 1094
781 TTTCAGGCTTAAATTTGATGATGATGATGATGATGATGATGATGATGAT 840
1095 CAATATCTGCAAGTGTATCATCTAAGCA-TTTTCAATCTCACTAAGTAACT 1153
841 CAATATCTGCAAGTGTATCATCTAAGCA-TTTTCAATCTCACTAAGTAACT 900
1154 TTACACATGCTTAAATTAACAAGAGTTGATTTGAGAGTCACTTGTGAATG 1213

Db. 901 TTGACACATCTTAATATC-AGCACTTCTCATTTGGAGTCACTTGAATGATCTT 959
OY 1214 CAAGGGAGACATATTGG 1232
Db 960 CCAGGGGGGCCCATATTGG 978

RESULT 2

AL574658/c 985 bp mRNA EST 16-FEB-2001
LOCUS AL574658 LFI_NFL006.PL2 Homo sapiens cDNA clone CSDDI065Y618 3

DEFINITION prime, mRNA, sequence.
ACCESSION AL574658
VERSION AL574658.1 GI:12935068

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 985)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..985
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDDI065Y618"
/clone_11b="LFI_NFL006.PL2"
/tissue_type="Placenta"

/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (41) 301 610 8371
Email : filiang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 261 a 212 c 162 g 322 t 28 others
ORIGIN

Query Match 70.5% Score 898.8; DB 10; Length 985;
Best Local Similarity 96.6%; Pred. No. 2.8e-209;
Matches 910; Conservative 26; Mismatches 4; Indels 2; Gaps 2;

OY 281 CCTTAAACCCCTATTAAACAGATGTGAAAAAGAAAACTCGCTATTTGCAATTGG 340
Db 953 CCTTAAACCCCTATTAAACAGATGTGAAAAAGAAAACTCGCTATTTGCAATTGG 895

OY 341 TTCCTGATTAAGGATATATCTGGAACATATGCTCCATCCTCAAGCTTGGAGACCA 400
Db 894 TTCCTGATTAAGGATATATCTGGAACATATGCTCCATCCTCAAGCTTGGAGACCA 835

OY 401 GGGCAATGATTAACATACCTGCTGTTGGTGCAATGA-CCCAATTGATGTGTGA 459
Db 834 GGGCAATGATTAACATACCTGCTGTTGGTGCAATGA-CCCAATTGATGTGTGA 775

OY 460 AATTGGAAGCAAGTATGTGCAAGAGTGAATTAATTGGCGTGAAGTCTAGCATATT 519
Db 774 AATTGGAAGCAAGTATGTGCAAGAGTGAATTAATTGGCGTGAAGTCTAGCATATT 715

OY 520 GGCATATGATTGACGAAGGGGAACCGACTGGAAGTCAATTGCCATTAAATGTGATGATCC 579
Db 714 GGCATATGATTGACGAAGGGGAACCGACTGGAAGTCAATTGCCATTAAATGTGATGATGCC 655

OY	580	TGTCACCACTAATTAATGATATCAATGATGTCAAAGCGTGAAACCTGGCTACTTGA	639
Db	654	TGATGCACCACATKATATATGATATCMAAGAATGTCAAAGCGTGAAACCTGGCTACTTGA	595
OY	640	AGCTACTGTGCAGCTGTTTAGAAGGTATTAAAGTTCCTCATGATGAAAACACGAAATATGAGTT	699
Db	594	AGCTACTGTGCAGCTGTTTAGAAGGTATTAAAGTTCCTCATGATGAAAACACGAAATATGAGTT	535
OY	700	TGCGTTTAATCCACAATTTAAAGATAAGGACTTTGGCCATTGATATTTATTAAGCACCTCA	759
Db	534	TGCGTTTAATCCACAATTTAAAGATAAGGACTTTGGCCATTGATATTTATTAAGCACCTCA	475
OY	760	TGACCATTTGGAAGACATAGTACATAAGAAAAACCAATGAAAAGAAATCAGTTCATGAA	819
Db	474	TGACCATTTGGAAGACATAGTACATAAGAAAAACCAATGAAAAGAAATCAGTTCATGAA	415
OY	820	TACAACTTTGTCGTGAGAGCCCCCTTCACAGTGTGATTCCTGATCCTGCCAGAGCATTTGGA	879
Db	414	TACAACTTTGTCGTGAGAGCCCCCBTCAAGTGTGATTCCTGATCCTGCCAGAGCATTTGGA	355
OY	880	TGCTTTACCAACACCCCTGTGAATCGCTCGACAGTACCAACACAGCGATTAAGTGTT	939
Db	354	TGCTTTACCAACACCCCTGTGAATCGCTCGACAGTACCAACACAGCGATTAAGTGTT	295
OY	940	CCATCCACCAAGAAAACCTATAGATTTCTCTGCAATAACACCTGATATTTGCTACATCTG	999
Db	294	CCATCCACCAAGAAAACCTATAGATTTCTCTGCAATAACACCTGATATTTGCTACATCTG	235
OY	1000	TTTCATCTGGATGATATTAGAGTAAAGATAGTAGCTTTTCAAGCTTTAATTTGTAGAAC	1059
Db	234	TTTCATCTGGATGATATTAGAGTAAAGATAGTAGCTTTTCAAGCTTTAATTTGTAGAAC	175
OY	1060	TCATCTACATTAAGTAATTTGCTGCTGCTACTAATTCATATTCACAGATGTTATTCATC	1119
Db	174	TCATCTACATTAAGTAATTTGCTGCTGCTACTAATTCATATTCACAGATGTTATTCATC	115
OY	1120	TAAAGCATTTTTCATATCTCAACTAGATTAACCTTTAGCACATGCTTAAATATCAAGCA	1179
Db	114	TAAAGCATTTTTCATATCTCAACTAGATTAACCTTTAGCACATGCTTAAATATCAAGCA	55
OY	1180	GTTGTCATTTGGAAAGTCACTGTGTAATGATGATGCAAGGGA	1221
Db	54	GTTGTCATTTGGAAAGTCACTGTGTAATGATGATGCAAGGGA	13
RESULT	3		
LOCUS	AK008575	1293 bp	mRNA
DEFINITION	Mus musculus adult male small intestine cDNA, RIKEN full-length		
ACCESSION	AK008575		
VERSION	AK008575.1	GI:12842842	
KEYWORDS	CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to mRNA, clone:lib.RIKEN full-length enriched mouse cDNA library		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus-		
TITLE	1 (bases 1 to 1293)		
JOURNAL	Carninci, P. and Hayashizaki, Y.		
MEDLINE	High-efficiency full-length cDNA cloning		
PUBMED	Methods in enzymology. 303, 19-44 (1999)		
REFERENCE	99279253		
AUTHORS	10349636		
TITLE	2 (bases 1 to 1293)		
JOURNAL	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,		
MEDLINE	Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
PUBMED	Normalization and subtraction of cap-trapper-selected cDNAs to		
	prepare full-length cDNA libraries for rapid discovery of new genes		
	Genome research. 10 (10), 1617-1630 (2000)		
	20499374		
	PUBMED	11042159	

```

REFERENCE
AUTHORS
3 (bases 1 to 1293)
Shibata,K., Itoh,M., Aizawa,K., Nagoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamanoto,R., Matsunoto,H., Sakaquchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohtsu,E., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multiscapillary sequencer
Genome research. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
20530913

TITLE
4 (bases 1 to 1293)

REFERENCE
AUTHORS
THE RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 1293)
Aachari,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arai,K., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Hori,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,T., Toyota,T., Yamamura,T., Yasunishi,A.,
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Shohiro-cho, Tsukuba-shi, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

COMMENT
FEATURES
SOURCE
Location/Qualifiers
1..1293
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1898257"
/db_xref="MGD:MGI:1915145"
/clone="2010317E03"
/sex="male"
/rnaseq_type="small intestine"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
109..978
/note="putative"
/codon_start=1
/protein_id="BAB25754.1"
/db_xref="GI:12842843"
/translation="MSCFSSEERAAPFTLEERYFLKNKKGQYISPFHDVPIYADKDFE
HMVVEPRMSNAKKEIATKDPLEIKQDVKGRKRVANLFPYGYIWNVGAIQWNE
DPGSHKHTGCCGNDPDIWCEIGSVCAREEIIKRVVLGILAMIDSGEDPMKVIATIN
VDPDDANAYKDISVERLKRYEIAATYDMPRPYRVPPGCKPNEFAFAPEKRPFAVD
IKSTHYRKALYTKTKTDGKICISCMNTTVSESPKCPDPAKAIYDALPPCESACSL
PTVDKWFHOKN"

BASE COUNT
354 a 288 c 336 g 315 t

ORIGIN
Query Match 65.6% Score 836.2 DB 12: Length 1293:
Best Local Similarity 82.4% Pred. No. 66-194:
Matches 1024: Conservative 0: Mismatches 203: Indels 16: Gaps 5:

```

[illegible]

QY	1173	CAAGCAGTGTCTATTGGACAGCACTTGATGATATATGCAAGGGGAGCAATATGG	1232
QY	1192	CAAAACCGTGGGACATCGAACATCTTGTGTCATGATGTGCAAGTGCACACCTATTGG	1251
OY	1233	ATGTATATCTTACCATATGTAGGAATTAATAATATTGGCG	1275
Db	1252	ATGTACATGTTTAC-TGCGTAGGAATAATAATATTACTG	1293

RESULT 4
AL519324
13-FEB-2001
DCM

DEFINITION AL519324 *U1-NTD₁-NCB2* gene
prime, mRNA sequence.

VERSION
KEYWORDS
EST.
human

Eukaryota; Metazoa; Chordata; Cnidaria; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Li, W.-B., Gruber, C., Josselyn, S., and Schmitt, J.
TITLE Full-length cDNA libraries and normalization

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

FEATURES	Location/Qualifiers
SOURCE	1. 826

```

/clone="CS0DA012YD02"
/clone lib="LTI_NFL011_NBC1"

```

```

/tissue_type="neuroblastoma cells
/lab_host="DH10B"
sequence 6: 1st strand

```

was primed with a Not I-enriched, double-stranded cDNA was digested with Not I and Eco RV sites of the pcMVSP

the normalized libraries. Library was constructed by technologies. Contact : Feng Liang Life Technologies.

Maryland 20850, USA Fax: (410) 301-630-3333
 Email: filangel@lifetech.com URL: www.lifetech.com

BASE COUNT	ORIGIN
231	1000

Query Match	Score	DB ID	Length
63.88;	813;	10;	020;
99.68;	Pred. No. 2.5e-188;		

47 CTGGTGGCTCTGTGGCAGCGGGCGGCGAGACTCCGGCACTATGAGGGCTTACAGACC 10

db 1 CAGGAGGACATCAGCGAATTCTTCCATAAAATGAGAAGGA 16

Db. 61 GAGGAGCGGCCGC - SCCTTCTCCTGGAGTACCGAGTATCCCAATTTCATCATG

Db 120 CAATATATATCTCCATTTCATGATATTCCAATTATGCGATAAGGATGCTTTTCATG 1

[illegible]


```

QY 287 AACCTATTAAACAGATGTGAAAAAGAAACCTGCTATGTGGCAATTGTCCG 346
Db 240 AACCTATTAAACAGATGTGAAAAAGAAACCTGCTATGTGGCAATTGTCCG 299
QY 347 TATTAAGATATATCTGGAACATGTGCTCCATCCCTCAGACTGGGAAGCCAGGCGAC 406
Db 300 TATTAAGATATATCTGGAACATGTGCTCCATCCCTCAGACTGGGAAGCCAGGCGAC 359
QY 407 AATGATTAACATCTGCTGTTGGTGACAAATGACCAATTTGATGTGTGAATTGGA 466
Db 360 AATGATTAACATCTGCTGTTGGTGACAAATGACCAATTTGATGTGTGAATTGGA 419
QY 467 AGCAAGTATGTCAGAGGTGAATTAATGGCTGAAGCTTCAGACATATGGCTATG 526
Db 420 AGCAAGTATGTCAGAGGTGAATTAATGGCTGAAGCTTCAGACATATGGCTATG 479
QY 527 ATTGACGAGGGGAAACCGACTGGAAGTCATTGCCATTATGTGATGATCCTGATGCA 586
Db 480 ATTGACGAGGGGAAACCGACTGGAAGTCATTGCCATTATGTGATGATCCTGATGCA 539
QY 587 GCCAATTATATGATATCATGATGATGCAAGGCTGAAGCTGCTACTATGAAGCTACT 646
Db 540 GCCAATTATATGATATCATGATGATGCAAGGCTGAAGCTGCTACTATGAAGCTACT 599
QY 647 GTGACTGCTTTAGAGATTAAGCTTCTGATGGAAGCAAGCAAAATGATTTGCTTT 706
Db 600 GTGACTGCTTTAGAGATTAAGCTTCTGATGGAAGCAAGCAAAATGATTTGCTTT 659
QY 707 AATGACGAATTTAAGATTAAGGACTTTGCCATGATATTTTAAAGCACTCATGACAT 766
Db 660 AATGACGAATTTAAGATTAAGGACTTTGCCATGATATTTTAAAGCACTCATGACAT 719
QY 767 TGGAAACATTAGTGAATGAAGAAAGATGGAAGAAATGCTGATGAATGACACT 826
Db 720 TGGAAACATTAGTGAATGAAGAAAGATGGAAGAAATGCTGATGAATGACACT 779
QY 827 TTGTCTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAGCAT 873
Db 780 TTGTCTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAGCAT 826

```

```

RESULT 5
LOCUS BG676397 932 bp mRNA EST 01-MAY-2001
DEFINITION 602622839F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:474797 5',
mRNA sequence.
ACCESSION BG676397
VERSION BG676397.1 GI:13907793
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 932)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
Plate: L1M10598 row: n column: 10
High quality sequence stop: 837.
location/Qualifiers
1..932
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

FEATURES

```

source
1..932
/location="taxon:9606"

```

```

/clone="IMAGE:474797"
/clone_id="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (rt1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-Sport6; Site: 1; NCI;
Site: 2; Sali; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 309 a 155 c 208 g 260 t
ORIGIN

```

```

Query Match 63.7%; Score 811.6; DB 11; Length 932;
Best Local Similarity 97.0%; Pred. No. 5,7e-188;
Matches 859; Conservative 0; Mismatches 24; Indels 3; Gaps 3;

QY 270 CTACAAAGACCCCTTTAAACCTTATTAACAGATGTGAAAAAGAAACCTGCTATG 329
Db 1 CTACAAAGACCCCTTTAAACCTTATTAACAGATGTGAAAAAGAAACCTGCTATG 60
QY 330 TTGGGAATTTGTTCCGTTAAAGATATATCTGGAACATGATGCTGATCCATCCAGACTT 389
Db 61 TTGGGAATTTGTTCCGTTAAAGATATATCTGGAACATGATGCTGATCCATCCAGACTT 120
QY 390 GGGAAAGACCCAGGACAGATGATTAACATGATGCTGCTGTTGTGTGACAAATGCAATTTG 449
Db 121 GGGAAAGACCCAGGACAGATGATTAACATGATGCTGCTGTTGTGTGACAAATGCAATTTG 179
QY 450 ATGTGTGTGAATTTGGAAGCAAGTATGTGCAAGGTGGAATTAATTTGGCTGAAAGTTC 509
Db 180 ATGTGTGTGAATTTGGAAGCAAGTATGTGCAAGGTGGAATTAATTTGGCTGAAAGTTC 239
QY 510 TAGCATATTGGCTATGATTAAGCAAGGGGAAACCGACTGGAAGTCAATTCCTATTAATG 569
Db 240 TAGCATATTGGCTATGATTAAGCAAGGGGAAACCGACTGGAAGTCAATTCCTATTAATG 299
QY 570 TGGATGATTCCTGATGACGCCAATTAATATGATTAATGATGCAAAAGGCTGAACCTG 629
Db 300 TGGATGATTCCTGATGACGCCAATTAATATGATTAATGATGCAAAAGGCTGAACCTG 359
QY 630 GCTACTTGAAGCTACTGTGACTGTTTGAAGCTATTAAGCTTCTGATGGAAGCAAG 689
Db 360 GCTACTTGAAGCTACTGTGACTGTTTGAAGCTATTAAGCTTCTGATGGAAGCAAG 419
QY 690 AAATGATTTGCTGCTTAATTCAGAAATTAAGTAAGGACTTTGCCATGATATTTATTA 749
Db 420 AAATGATTTGCTGCTTAATTCAGAAATTAAGTAAGGACTTTGCCATGATATTTATTA 479
QY 750 AAAGCCTCATGACCATTTGAAGCAATTAAGTATGAAGAAAGCAAGTAAGCAATCA 809
Db 480 AAAGCCTCATGACCATTTGAAGCAATTAAGTATGAAGAAAGCAAGTAAGCAATCA 539
QY 810 GTTCATGATTAACATTTGTCTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAG 869
Db 540 GTTCATGATTAACATTTGTCTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAG 599
QY 870 CCATTGTGATGCTTTACACACACCGCTGTAATGCTGCTGCAAGTACCAAGACGCTGG 929
Db 600 CCATTGTGATGCTTTACACACACCGCTGTAATGCTGCTGCAAGTACCAAGACGCTGG 659
QY 930 ATAGGTGTTCCATCACAGAAAACTAATGAGATTTCTGTAATACAGCTGATATTG 989
Db 660 ATAGGTGTTCCATCACAGAAAACTAATGAGATTTCTGTAATACAGCTGATATTG 719
QY 990 CTACATGCTGTTCACTGAGATTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1048
Db 720 GTACATGCTGTTCACTGAGATTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 779
QY 1049 ATTGTGAACTCATTAACATAAGTAAATTCCTGCTGACTATTCATTAATCAAGTAA 1108
Db 780 ATTGTGAACTCATTAACATAAGTAAATTCCTGCTGACTATTCATTAATCAAGTAA 839
QY 1109 TGTATATCATTAAGCAATTTTCAATATCTCACTAAGATTAATCTTT 1154

```


Db 244 GGTATGTTGGCAATTTGTTCCGGTATTAAGATATATCGAATAGTGGCCATCCCTC 303
OY 384 AGACTTGGAGAGACCCAGGACAAATGATTAACATCTAGCTGCTGTGTGTCATGAC 443
Db 304 AGACTTGGAGAGACCCAGGACAAATGATTAACATCTAGCTGCTGTGTGTCATGAC 363
OY 444 CAATGATGTTGTTGAATTTGGAAAGGATGTCGAAGAGCTGAATTAATTTGGCGTGA 503
Db 364 CAATGATGTTGTTGAATTTGGAAAGGATGTCGAAGAGCTGAATTAATTTGGCGTGA 423
OY 504 AAGTCTAGGAGCAATTTGGCTATGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 563
Db 424 AAGTCTAGGAGCAATTTGGCTATGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 483
OY 564 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 623
Db 484 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 543
OY 624 AACCTGGCTACTTAGAAGCTAGTGAAGCTGAGCTGTTAGAAAGTATTAAGTCTGATGAA 683
Db 544 AACCTGGCTACTTAGAAGCTAGTGAAGCTGAGCTGTTAGAAAGTATTAAGTCTGATGAA 603
OY 684 AACCGAAAATAGTGTGCTTTAATGAGATTT-AAAAGATAGAGCTTTGCCATGAT 742
Db 604 AACCGAAAATAGTGTGCTTTAATGAGATTTAAGATTAAGATTAAGATTAAGATTAAG 663
OY 743 ATTATTAAGACATCATGACATGACATGACATGACATGACATGACATGACATGACATG 801
Db 664 ATTATTAAGACATCATGACATGACATGACATGACATGACATGACATGACATGACATG 723
OY 802 AGCAATCAAGTTCATGATATACATCACTTTGTTGAGAGAGCCCTTCAAGTGTG-851
Db 724 AGCAATCAAGTTCATGATATACATCACTTTGTTGAGAGAGCCCTTCAAGTGTG 773

RESULT 9
B1258591 771 bp mRNA EST 17-JUL-2001
LOCUS 602969519P1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5109100 5'
DEFINITION mRNA sequence.
ACCESSION B1258591
VERSION B1258591.1 GI:14815097
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 771)
NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
AUTHORS Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM11264 row: e column: 05
High quality sequence stop: 749.
Location/Qualifiers
1..771
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5109100"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-Sport6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectional. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life

BASE COUNT 264 a 132 c 155 g 220 t
ORIGIN
Technology: "Best Local Similarity 99.3%; Pred. No. 5,7e-11; Matches 745; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Query Match 58.2%; Score 742; DB 11; Length 771;
Best Local Similarity 99.3%; Pred. No. 5,7e-11; Matches 745; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 521 GGTATGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 580
Db 1 GGTATGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 60
OY 581 GATGAGGCAATTTAATGATATCAATGATGATGATGATGATGATGATGATGATGATGAT 640
Db 61 GATGAGGCAATTTAATGATATCAATGATGATGATGATGATGATGATGATGATGATGAT 120
OY 641 GCTACTGAGAGCTGTTAGAAAGTATTAAGTCTGATGATGATGATGATGATGATGATG 700
Db 121 GCTACTGAGAGCTGTTAGAAAGTATTAAGTCTGATGATGATGATGATGATGATGATG 180
OY 701 GCGTTAATGAGAGATTTAAGATTAAGAGCTTTGCCATTTAATTAAGAGCACTGAT 760
Db 181 GCGTTAATGAGAGATTTAAGATTAAGAGCTTTGCCATTTAATTAAGAGCACTGAT 240
OY 761 GACCAATGAGAGCAATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 820
Db 241 GACCAATGAGAGCAATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 300
OY 821 ACAATTTGCTGAGAGAGCCCTTCAAGTGTGATGATGATGATGATGATGATGATGATGAT 880
Db 301 ACAATTTGCTGAGAGAGCCCTTCAAGTGTGATGATGATGATGATGATGATGATGATGAT 360
OY 881 GCTTTACCAACACCTGTAAGTGTGCTGACAGTACAGACAGAGAGAGAGAGAGAGAGAG 940
Db 361 GCTTTACCAACACCTGTAAGTGTGCTGACAGTACAGAGAGAGAGAGAGAGAGAGAGAG 420
OY 941 CATCACAGAAAACAACTAATGAGATTTCTGTAATGATGATGATGATGATGATGATGATG 1000
Db 421 CATCACAGAAAACAACTAATGAGATTTCTGTAATGATGATGATGATGATGATGATGATG 480
OY 1001 TCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1060
Db 481 TCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
OY 1061 CATCTAAGTAAGTAATTTCTGCTGATGATGATGATGATGATGATGATGATGATGATG 1120
Db 541 CATCTAAGTAAGTAATTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 600
OY 1121 AAGCAATTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1180
Db 601 AAGCAATTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
OY 1181 TTGCTATTTGGAAGCACTTTGTAATGATGATGATGATGATGATGATGATGATGATGAT 1240
Db 661 TTGCTATTTGGAAGCACTTTGTAATGATGATGATGATGATGATGATGATGATGATGAT 720
OY 1241 GTTACATATGTTAGAAATTAATTTT 1270
Db 721 GTTACATATGTTAGAAATTAATTTT 750

RESULT 10
Bg713976 772 bp mRNA EST 08-MAY-2001
LOCUS 602674374P1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4796875 5'
DEFINITION mRNA sequence.
ACCESSION Bg713976
VERSION Bg713976.1 GI:13992907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 772)	Mammalia: Eutheria: Primates; Catarrhini; Hominoidea; Homo.	NIH-MGC http://mgc.nci.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC).	unpublished (1999)
Contact:	Robert Strausberg, Ph.D.			

BASE COUNT	235 a	146 c	197 g	194 t
ORIGIN				

Query Match	58.0%	Score 739.8	DB 11	Length 772
Best Local Similarity	99.5%	Pred. No. 2e-170		
Matches 763	Conservative	0	Mismatches 2	Indels 2

QY	28	AGTGGGCGCCGCGTGTGCGGGGTGTGGGCTGTGGGACGGGGCGGACAGACATCCGGGAC	87
Db	6	AGTGGGCGCCCGCTGTCGGGGCTGTGGGCTGTGGGACGGGGCGGACGCTCCGGGAC	65
QY	88	TATGAGCGGCTTCAGCACCGAGAGAGCGCGCGCCCTCTCTCCCTGGATACGACTCTT	147
Db	66	TATGAGCGGCTTCACACCGAGAGAGCGCGCGCCCTCTCTCCCTGGATACGACTCTT	125
QY	148	CCTCAAAATATGAAAGAGACATATATATCTTCATTTCTATTTCCAAATTATGACA	207
Db	126	CCTCAAAATATGAAAGAGACATATATATCTTCATTTCTATGATATTTCCAAATTATGACA	185
QY	208	TAAGATGTGTTTCATCATGTGTGAAGTACACGCGTGTCTAATGCAAAATGGAGAT	267
Db	186	TAAGATGTGTGTTTCATCATGTGTGAAGTACACGCGTGTCTAATGCAAAATGGAGAT	245
QY	268	TGCTACAAAGACCCCTTTAAACCCTATTATTAACACAGATGTGAAAAAGAAAACTTGCTTA	327
Db	246	TGCTACAAAGACCCCTTTAAACCCTATTATTAACACAGATGTGAAAAAGAAAACTTGCTTA	305
QY	328	TGTTGCCAATTTGTTCCCGTATAAAGGATATATCTGGAACTATGGTGGCATCCCTAGAC	387
Db	306	TGTTGCCAATTTGTTCCCGTATAAAGGATATATCTGGAACTATGGTGGCATCCCTAGAC	365
QY	388	TTGGGAACACCCAGGGACCAATGATAAACATCTGGCTTGTGGTGACAATGACCCAAAT	447
Db	366	TTGGGAACACCCAGGGACCAATGATAAACATCTGGCTTGTGGTGACAATGACCCAAAT	425
QY	448	TGATGTGTGTGAATTTGGAAGCAAGGTATGTGCAAGAGGTGAATAATTTGGCGTGAAGT	507
Db	426	TGATGTGTGTGAATTTGGAAGCAAGGTATGTGCAAGAGGTGAATAATTTGGCGTGAAGT	485

Oy	508	TCTAGCATTATTTGGCTATGATGACGAAGGGAAACCGACTGGAAGTATATGGCATTAA	567
Db	486	TCTAGCATTATTTGGCTATGATGACGAAGGGAAACCGACTGGAAGTATATGGCATTAA	545
Oy	568	TGTGATGATATCCTGATGCGCCCAATTATATATCATATGATGTCTAAAGGCTGAAAC	627
Db	546	TGTGATGATATCCTGATGCGCCCAATTATATATCATGATGTCTAAACGGCTGAAAC	605
Oy	628	TGGCTACTTAGAAGGTACGTGGACTGGTTTGAAGGTATTAAGTGTCCGATGGAAGAAC	687
Db	606	TGGCTACTTAGAAGGTACGTGGACTGGTTTGAAGGTATTAAGTGTCCGATGGAAGAAC	665
Oy	688	AGAAATATC-AGTTGCGTTTAAATGCAGAAATTTAAAGATPAGACATTTGCCATTTGA-TATT	745
Db	666	AGAAATATGAGTTTCCGTTTAAATGCAGAAATTTAAAGATPAGACATTTGCCATTTGACTATT	725
Oy	746	ATTAAAGACATCATGACCATTTGGAAAGCATTTAGTGACTTAAAGAAAC	792
Db	726	ATTAAAGACATCATGACCATTTGGAAAGCATTTAGTGACTTAAAGAAC	772

RESULT	11
B1259308	
LOCUS	
DEFINITION	B1259308 878 bp mRNA EST 17-JUL-2001 60297254.F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5111563 5' ,
ACCESSION	mRNA sequence.
VERSION	B1259308
KEYWORDS	B1259308.1 GI:14816515
SOURCE	EST .
ORGANISM	human;
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 878)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

High quality sequence, stop:

BASE COUNT	295 a	141 c	192 g	250 t
ORIGIN				

Qy 423 GCCTGTTGTGCTGACCAATGACCCATTGATGTGTGAAATTGGAAAGCAAGTATGTGCAA 482
|||||
Db 1 GCCTGTTGTGCTGAC-ATGACCCATTGATGTGTGTGAAATTGGAAAGCAAGTATGTGCAA 59
|||||


```

/db_xref="MGI:1915145"
/clone="9130233620"
/sex="male"
/issue_type="cecum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
112..591
CDS
/note="putative"
/codon_start=1
/protein_id="BAB32057.1"
/db_xref="GI:12860833"
/translation="MSGFSSERRAPFLLEYRVLKNEKGOYISPHVDPIYADKDF
HMVYEPKSNKMKIATKPLNPIKOVKKKLYVANLPYKRTINNYGAIPQWE
DPGSHDKAGCGDNDPIDVCEIGSKVARSITIKVKVLIAMIDEGEDRLRGCH"
1291..1296
polyA_signal
/note="putative"
1308
polyA_site
/note="putative"
293 c 338 g 319 t
BASE COUNT 358 a 293 c 338 g 319 t
ORIGIN

Query Match 57.3% Score 730.2; DB 12; Length 1308;
Best Local Similarity 80.4%; Pred. No. 5e-168;
Matches 1009; Conservative 0; Mismatches 218; Indels 28; Gaps 12;

QY 40 GTGCGGGCTGTGCTCTGTGCGAGCGGCGGAGACTCGGCACTATGAGCGGCTT 99
DB 63 GTGTGTTGTGTGCTGTGCGAGCGGCGGCGGAGACTCTCGACACCATGAGCGGCTT 122
QY 100 CAGCACCGAGAGCGCGCGCGGCGGCTTCCCTGAGTACCGAGCTTCTCAAAATGA 159
DB 123 CAGCAGCGAGAGCGCGCGCGGCGGCTTCCCTGAGTACCGAGCTTCTCAAAATGA 182
QY 160 GAAAGGACATATATTCATTCATATTCATATTCATATTCATATTCATATTCAT 219
DB 183 GAAAGGACATATTCATTCATATTCATATTCATATTCATATTCATATTCATATTC 242
QY 220 TCACATGTAGTGTGAATACCAAGCTGTCTATGCAAAATGAGATGTGTACAAGA 279
DB 243 CCACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 302
QY 280 CCCTTTAAACCTATTTAAACAAGATGTGAAGAAAGGAAAGGAAAGGAAAGGAA 339
DB 303 CCCTTTAAACCTATTTAAACAAGATGTGAAGAAAGGAAAGGAAAGGAAAGGAA 362
QY 340 GTTCCCGTAAAGATATATCTGTGCAAGATGTGTGCAATGCTGCAATGCTGCA 399
DB 363 GTTCCCGTAAAGATATATTCATATTCATATTCATATTCATATTCATATTCAT 422
QY 400 AGGCGACATGTAAACATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 459
DB 423 AGGCGACATGTAAACATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 482
QY 460 AATGTGACCAAGTATGTGCAAGAGTGAATATTTGGCGTGAAGTGTCTAGCAATT 519
DB 483 AATGTGACCAAGTATGTGCAAGAGTGAATATTTGGCGTGAAGTGTCTAGCAATT 542
QY 520 GCGTATGATGTGCAAGAGGAA-ACCGAGTGAAGTGTGTGTGTGTGTGTGTGTGT 576
DB 543 GCGTATGATGTGCAAGAGGAA-ACCGAGTGAAGTGTGTGTGTGTGTGTGTGTGT 602
QY 577 TCGTGTATGC--AGCCAATTAATATGATA-TCAATGATGTGCAAGCGCTGAAA---CGTGG 630
DB 603 TCGGAGCGCGACCCCTTAATTAAGATATTTCTGTGATGTGCAAGCGGCTTAAACCCGGG 662
QY 631 CTACTTAAAGTACT--GTGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 687
DB 663 CTATTATGAAGTACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 722
QY 688 AAAAAATGTTGCTTGTATGCAAGATTTAAAGATGAGATGAGATGAGATGAGATGAG 747
DB 723 AAAAAATGTTGCTTGTATGCAAGATTTAAACCAAGGACTTGTGAGATGAGATGAG 782

```

```

QY 748 TAAAGCACTCATGACCATTGTGAAGACATTAGTACTAGAAAAAGAAAGCAAT 807
DB 783 TAAAGCACTCATGACCATTATTTGAAAGCAATTAGTACTAGAAAAAGAAAGCAAT 842
QY 808 CAGTTCATGATATCAACACTTGTCTGAGAGCCCTCAAGTGTGATTCCTGATCTGCGAC 867
DB 843 CAGTTCATGATATCAACACTTGTCTGAGAGCCCTCAAGTGTGATTCCTGATCTGCGAC 902
QY 868 AGCCATTGTGATGCTTTTACACACACCCCTGTGATGCTGCTGACAGTACCAAGACGT 927
DB 903 AGCCATTGTGATGCTTTTACACACACCCCTGTGATGCTGCTGACAGTACCAAGACGT 962
QY 928 GATATAGTGTTCATCACCAGAAAAAATAAGATATTTCTGTGAAATACAGCTGATAT 987
DB 963 GATATAGTGTTCATCACCAGAAAAAATAAGATATTTCTGTGAAATACAGCTGATAT 1022
QY 988 TGTATCATGCTGTTCATCTGTGATGATGTATGAAAGTAAAGTATGAGCTTTCAAGCTTTA 1047
DB 1023 GCGCGCATCTCTGTTCACCTGGAAGCATTTAGATCAAAAGA-----TTTTCAAGCTCTA 1076
QY 1048 AATTGTAGACTCATCTACTAAGTAAATTTGCTGTGACTAATTCATATATCTACAGA 1107
DB 1077 AATTGTGTAGCTCACCAGAA-TGAAGTACACTCTGCTGTGACTGTCTACTACTTACAGAA 1135
QY 1108 ATGTTATCCATCTAAGCAATTTT-----TCATATCTCACTAATGATTAATTTAGCAC 1160
DB 1136 CTGTGTGAGAGGTGTAGCCATCTAGACACCTCATCTCATCAACAGACGCTTTATTA 1194
QY 1161 ATGCTTAATATCAAAAGCACTGTCTATTTGGAAGTACACTTGTGAATATGATGTCAAGGGG 1220
DB 1195 AACCTTAGCTCTCAAAAGCGGTGGGACATGGAAGTATTTGTCAATGATGTCACTGTGTC 1254
QY 1221 AGCACATTTGATGTATATGTTATACATATGTTAGGAATAATATTTTGTGTG 1275
DB 1255 GACACCTATTTGATGTATGATGTTTAC-TGTGTAGGAATAATATTTTACCAG 1308

```

```

RESULT 13
LOCUS BG574468
DEFINITION 602596421F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4705253 5',
ACCESSION BG574468
VERSION BG574468.1 GI:13582121
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 915)
AUTHORS NIH-MGC http://mgi.mcl.nih.gov/
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM10571 row: 3 column: 06
High quality sequence stop: 830.
Location/Qualifiers
1..915
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4705253"
/clone_lib="NIH_MGC_87"
/issue_type="mammalian adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"

```

FEATURES

source

/note="Organ: breast; Vector: PCMV-SPORT6; Site:1: NOTI;
 Site:2: Sall; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.383 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH-MGC Library."

BASE COUNT 269 a 185 c 231 g 230 t
 ORIGIN

Query Match 56.5%; Score 721; DB 11; Length 915;
 Best Local Similarity 94.1%; Pred. No. 8.1e-166;
 Matches 850; Conservative 0; Mismatches 35; Indels 18; Gaps 9;

29 GTGGGGCCCGCTGGGGGCTGTGCTCTGTGACAGCGGGCGGAGAGATCCGGCACT 88
 1 GTGGGGCCCGCTGGGGGCTGTGCTCTGTGACAGCGGGCGGAGAGATCCGGCACT 60

89 ATGAGCGGCTTCAGACACGAGAGCGCGCGCCCTTCCTCCGTGAGTACGAGTCTTC 148
 61 ATGAGCGGCTTCAGACACGAGAGCGCGCGCCCTTCCTCCGTGAGTACGAGTCTTC 120

149 CTCAGAAATGAGAAAGAGCAATATATATCTCATTTTCCATTTTATGAGAT 208
 121 CTCAGAAATGAGAAAGAGCAATATATATCTCATTTTCCATTTTATGAGAT 180

209 AAGGATGTGTTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 268
 181 AAGGATGTGTTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 240

269 GCTACAAAGAGCCCTTAAACCATTAACCAAGATGGAAGAAAGAAATTCGCTAT 328
 241 GCTACAAAGAGCCCTTAAACCATTAACCAAGATGGAAGAAAGAAATTCGCTAT 300

329 GTTGGCAATTTGTTCCCTATTAAGATATATGGAATCTATGTCATTCCTCAACT 388
 301 GTTGGCAATTTGTTCCCTATTAAGATATATGGAATCTATGTCATTCCTCAACT 360

389 TGGGAAGACCCAGGCGCAATATATATATATATATATATATATATATATATAT 448
 361 TGGGAAGACCCAGGCGCAATATATATATATATATATATATATATATATATAT 420

449 GATGTGTGGAATTTGGAAGCAAGGTATGTGCAAGAGGTGAATATATGCGGTAACT 508
 421 GATGTGTGGAATTTGGAAGCAAGGTATGTGCAAGAGGTGAATATATGCGGTAACT 480

509 CTAGGCAATTTGCTATGATTTGACAGAGGGGAAACCGACTGGAAGTCAATTTAT 568
 481 CTAGGCAATTTGCTATGATTTGACAGAGGGGAAACCGACTGGAAGTCAATTTAT 540

569 GTGGATGATCCCTGATGACCAATTAATATATATATATATATATATATATATAT 628
 541 GTGGATGATCCCTGATGACCAATTAATATATATATATATATATATATATATAT 600

629 GGCTACTTGAAGCTACTGTGACTGTTTAA--AAGTATTAAGTTCCTGATGGAAC 686
 601 GGCTACTTGAAGCTACTGTGACTGTTTAA--AAGTATTAAGTTCCTGATGGAAC 660

687 CAGAAAATGAGTTGGCTTTAATGCAAGATTTAAGATAGAGCTTGGCATTTGAT 744
 661 CAGAAAATGAGTTGGCTTTAATGCAAGATTTAAGATAGAGCTTGGCATTTGAT 720

745 TATTTAAAGACTATGACCATTTGAAAGCATTTAGACTATAGAAA-----ACGAATG 798
 721 TATTTAAAGACTATGACCATTTGAAAGCATTTAGACTATAGAAA-----ACGAATG 779

799 AAAAGATATGATTTGATGAT--ACAATTTTCTGAGAGCGCCCTTCAAGT--GTGATCC 855
 780 AAAAGATATGATTTGATGAT--ACAATTTTCTGAGAGCGCCCTTCAAGT--GTGATCC 839

856 TG--ATCTGCGACAGCATTTGGA--TGCTTTACCAAC--ACCCTGTGAATTCGCTGCA 911
 840 TG--ATCTGCGACAGCATTTGGA--TGCTTTACCAAC--ACCCTGTGAATTCGCTGCA 899

OR 912 CAG 914
 Db 900 CAG 902

RESULT 14
 LOCUS BG708966
 DEFINITION 602675112F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4797723 5',
 mRNA sequence.
 ACCESSION BG708966
 VERSION BG708966.1 GI:13986834
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 845)
 NIH-MGC <http://mhc.ncl.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshitsuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM10684 row: 9 column: 04
 High quality sequence stop: 712.

FEATURES
 source location/Qualifiers
 1..845
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4797723"
 /clone_lib="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: Brain; Vector: pBluescript (modified
 pBluescript KS+); Site:1: BamHI; Site:2: Sall; XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTATTTTATTTTATTTTATTTT-3',
 size-selected for average insert size 2.3 kb and
 normalized to 10^5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 246 a 160 c 232 g 207 t
 ORIGIN

Query Match 56.5%; Score 720.4; DB 11; Length 845;
 Best Local Similarity 96.0%; Pred. No. 1.1e-165;
 Matches 782; Conservative 0; Mismatches 27; Indels 6; Gaps 4;

5 AGTTTGGGCT 64
 24 AGGCTTGGGCT 83

65 GGGGGGGGCGAGAGCTCCGGCACTATGAGCGGCTTCAGACACGAGAGCGCGCGCC 124
 84 GGGGGGGGCGAGAGCTCCGGCACTATGAGCGGCTTCAGACACGAGAGCGCGCGCC 143

125 TTCTCCCTGAGACTACCGAGTCTTCTCAAAAATGGAAGAGACATATATCTCCATTT 184
 144 TTCTCCCTGAGACTACCGAGTCTTCTCAAAAATGGAAGAGACATATATCTCCATTT 203

185 CATGATATTCATTTATGACAGATAGAGATGTGTTTCATGATGATGATGATGATGAT 244
 204 CATGATATTCATTTATGACAGATAGAGATGTGTTTCATGATGATGATGATGATGAT 263

Mon Apr 1 07:59:58 2002

us-09-415-540-2.std.rst

Page 14

Search completed: March 29, 2002, 19:50:17
Job time: 3420 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 19:50:25 ; Search time 149.95 seconds

(without alignments)
9134.216 Million cell updates/sec

Title: US-09-415-540-2

Perfect score: 1275
Sequence: 1 CAAGAGCTTGGCGCTCTCT.....GAATTAATAATTTTCCTG 1275

Scoring table: Oligo_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estlin: *
4: em_estlpl: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: qb_estl: *
11: qb_est2: *
12: qb_hic: *
13: qb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rtd: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	751	58.9	826	10	AL519324
2	709	55.6	714	2	BG700549
3	656	51.5	771	11	BI258591
4	636	49.9	773	11	BG715195
5	635	49.8	905	11	BG027821
6	632	49.6	915	11	BG574468
7	628	49.3	628	11	BE887465
8	625	49.0	660	10	AM009649
9	618	48.5	772	11	BG713976
10	607	47.6	955	11	BF791373
11	599	47.0	602	11	BG504546
12	595	46.7	596	11	BG574238

C	13	592	46.4	998	11	BF966851	BF966851	602286503
C	14	586	46.0	932	11	BG676397	BG676397	602622839
C	15	584	45.8	584	10	AT1719667	AT1719667	at412.x
C	16	581	45.6	996	11	BG674343	BG674343	602620303
C	17	579	45.4	882	11	BF966192	BF966192	602286503
C	18	577	45.3	711	11	BG777143	BG777143	602664359
C	19	575	45.1	575	10	AM262869	AM262869	xg96e02.x
C	20	572	44.9	714	11	BG702955	BG702955	602684915
C	21	570	44.7	769	11	BG702317	BG702317	602683563
C	22	567	44.5	567	11	BG180287	BG180287	602331122
C	23	567	44.5	702	10	AM840924	AM840924	RC1-CN000
C	24	562	44.1	1076	11	BG179769	BG179769	602328856
C	25	557	43.7	585	10	AM873109	AM873109	hg22h11.x
C	26	555	43.5	589	10	AM659221	AM659221	hc79e09.x
C	27	552	43.3	592	11	BE867111	BE867111	601442951
C	28	544	42.7	762	11	BG721093	BG721093	602692932
C	29	543	42.6	639	10	AA694204	AA694204	z12c12.s
C	30	541	42.4	564	10	AA493425	AA493425	ne48h03.s
C	31	539	42.3	750	11	BG704401	BG704401	602687468
C	32	535	42.0	593	11	BG714352	BG714352	602669861
C	33	532	41.7	914	11	BF033412	BF033412	601457932
C	34	528	41.4	631	10	AM518147	AM518147	xx81f10.x
C	35	523	41.0	570	10	AT215898	AT215898	qm35h09.x
C	36	517	40.5	690	10	AV706813	AV706813	AV706813
C	37	516	40.5	775	11	BG036446	BG036446	602326583
C	38	515	40.4	794	10	BE379621	BE379621	601159361
C	39	510	40.0	617	10	AT1818185	AT1818185	wk42c03.x
C	40	504	39.5	940	11	BG254456	BG254456	602369062
C	41	503	39.5	503	11	BE887538	BE887538	601508191
C	42	503	39.5	585	10	BE042554	BE042554	hc25e08.x
C	43	503	39.5	624	11	BG777846	BG777846	602664958
C	44	503	39.5	804	11	BI260339	BI260339	602696372
C	45	502	39.4	624	10	AT922084	AT922084	wm8db03.x

ALIGNMENTS

RESULT 1
AL519324 826 bp mRNA EST 13-FEB-2001
LOCUS AL519324 LTI_NFL011.NBC1 Homo sapiens cDNA clone CSODAO12YD02 5
DEFINITION AL519324 LTI_NFL011.NBC1 Homo sapiens cDNA clone CSODAO12YD02 5
prime, mRNA sequence.
ACCESSION AL519324
VERSION AL519324.1 GI:12782817
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 826)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source
1. 826
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODAO12YD02"
/clone_lib="LTI_NFL011.NBC1"
/sex="male"
/tissue="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5109100"
/clone_1lb="NH MGC 12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/notes="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT Average insert size 1.4 Kb. Library prepared by Ligo

```

BASE COUNT	1	264 a	132 c	155 g	220 t
ORIGIN	Technologies."				
Query Match	51.5%; Score 656; DB 11; Length 771;				
Best Local Similarity	99.9%; Pred. No. 0;				
Matches 706; Conservative	0; Mismatches 1; Indels 0; Gaps 0;				
OY	521	GCTATGATTGACGAAGGGGAAACCGACACTGGGAAAGTCATTGCCATTAAATGTGATGATCCT	580		
DB	1	GCTATGATTGACGAAGGGGAAACCGACACTGGGAAAGTCATTGCCATTAAATGTGATGATCCT	60		
OY	581	GATGAGGCCAATTATTAATCATATGCAATGTGCAAAAGGCTGAACCTGGCTACTTAGAA	640		
DB	61	GATGAGGCCAATTATTAATCATATGCAATGTGCAAAAGGCTGAACCTGGCTACTTAGAA	120		
OY	641	GCTACTGTGACACTGGTTTGAAGATATAGTATTCCTGATGAGGAAACACGAAATAGATT	700		
DB	121	GCTACTGTGACACTGGTTTGAAGATATAGTATTCCTGATGAGGAAACACGAAATAGATT	180		
OY	701	GGGTTTAATGACGAATTTAAAGATAAGGACTTTGCCATTGATATTTAAAGCACTCAT	760		
DB	181	GGGTTTAATGACGAATTTAAAGATAAGGACTTTGCCATTGATATTTAAAGCACTCAT	240		
OY	761	GACCAATTGGAAGCAATTACTGACTAAGAAACGAATGGAAGAAAGGATCAGTTGCATGAT	820		
DB	241	GACCAATTGGAAGCAATTACTGACTAAGAAACGAATGGAAGAAAGGATCAGTTGCATGAT	300		
OY	821	ACAACCTTGCTGACGAGGCCCTTCAAGTGTGATGCCGATGTCGCCAGACCATTTGGGAT	880		
DB	301	ACAACCTTGCTGACGAGGCCCTTCAAGTGTGATGCCGATGTCGCCAGACCATTTGGGAT	360		
OY	881	GCTTTACGACCAACCCCTGTGAATCTGCTGACACAGTACCAACAGAGCTGATTAAGTGGTTC	940		
DB	361	GCTTTACGACCAACCCCTGTGAATCTGCTGACACAGTACCAACAGAGCTGATTAAGTGGTTC	420		
OY	941	CATCACACGAAAACTAATGACATTTCTTGGAAATCAAGCTGATATTCTCTACATCTGT	1000		
DB	421	CATCACACGAAAACTAATGACATTTCTTGGAAATCAAGCTGATATTCTCTACATCTGT	480		
OY	1001	TCATCGGATGATTAAGTAAGTAAGTAAGTACTAGCTTTTCAAGCTTTAAATTTGTAGAACT	1060		
DB	481	TCATCGGATGATTAAGTAAGTAAGTAAGTACTAGCTTTTCAAGCTTTAAATTTGTAGAACT	540		
OY	1061	CATCTAACAATAAGTAATTCCTGCTGTGACTAATCCATATACTCAGAAATGTTATCCATCT	1120		
DB	541	CATCTAACAATAAGTAATTCCTGCTGTGACTAATCCATATACTCAGAAATGTTATCCATCT	600		
OY	1121	AAAGCATTTTATATCTCACTAAGATTAACCTTTTAGCAATGCTTAAATATCAAAAGCAG	1180		
DB	601	AAAGCATTTTATATCTCACTAAGATTAACCTTTTAGCAATGCTTAAATATCAAAAGCAG	660		
OY	1181	TTTGCTATTGGAAGTCATCTTGGAATAGATGTGCAAGGGGAGACACAT	1227		
DB	661	TTTGCTATTGGAAGTCATCTTGGAATAGATGTGCAAGGGGAGACACAT	707		
RESULT	4				
LOCUS	BT715195	773 bp	mRNA	EST	08-MAY-2001
DEFINITION	602675655F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4798285 5'				
ACCESSION	BT715195				
VERSION	BT715195.1	GI:13994274			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 773)				
	NIH-MGC hclup://mhc.nci.nih.gov/.				
	National Institutes of Health, Mammalian Gene Collection (MGC)				

OY 747 TTAAGACATGACGACATTGAAAGCATTAGTACTAGAGAAAAGCAATGAGAAAGGAA 806
|||||
DB 241 TTAAGACATGACGACATTGAAAGCATTAGTACTAGAGAAAAGCAATGAGAAAGGAA 300
OY 807 TCAGTGCATGATGATGACATTTGCTGTGAGAGCCCTTCAAGTGTATCTGATGCTGCA 866
|||||
DB 301 TCAGTGCATGATGATGACATTTGCTGTGAGAGCCCTTCAAGTGTATCTGATGCTGCA 360
OY 867 GAGCAGATGATGATGACATTTGCTGTGAGAGCCCTTCAAGTGTATCTGATGCTGCA 926
|||||
DB 361 GAGCAGATGATGATGACATTTGCTGTGAGAGCCCTTCAAGTGTATCTGATGCTGCA 420
OY 927 TGATGATGATGATGACATTTGCTGTGAGAGCCCTTCAAGTGTATCTGATGCTGCA 986
|||||
DB 421 TGATGATGATGATGACATTTGCTGTGAGAGCCCTTCAAGTGTATCTGATGCTGCA 480
OY 987 TTGCTACATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1046
|||||
DB 481 TTGCTACATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
OY 1047 AAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1106
|||||
DB 541 AAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
OY 1107 AATGTTATCATCTTAAGCATTGTTTCATATCTCA 1141
|||||
DB 601 AATGTTATCATCTTAAGCATTGTTTCATATCTCA 635

RESULT 6
BG574468 915 bp mRNA EST 10-APR-2001
LOCUS 602596421F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4705253 5',
DEFINITION mRNA sequence.
ACCESSION BG574468
VERSION BG574468.1 GI:13582121
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 915)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LRAM10571 row: j column: 06
High quality sequence stop: 830.

FEATURES
Source
Location/Qualifiers
1..915

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4705253"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 269 a 185 c 231 g 230 t
ORIGIN

Query Match 49.6%; Score 632; DB 11; Length 915;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 GTGGGCGCGCGGTGGGCGGTGGGCGGTGGGCGGTGGGCGGTGGGCGGTGGGCGGTGG 88
|||||
DB 1 GTGGGCGCGCGGTGGGCGGTGGGCGGTGGGCGGTGGGCGGTGGGCGGTGGGCGGTGG 60
OY 89 ATGAGCGGCTTCAGACCGAGAGGAGGCGCGCGGTGGGCGGTGGGCGGTGGGCGGTGG 148
|||||
DB 61 ATGAGCGGCTTCAGACCGAGAGGAGGCGCGCGGTGGGCGGTGGGCGGTGGGCGGTGG 120
OY 149 CTCAAAATGAGAAAGCAAT 208
|||||
DB 121 CTCAAAATGAGAAAGCAAT 180
OY 209 AAGATGCTGTTACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 268
|||||
DB 181 AAGATGCTGTTACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
OY 269 GCTAAGAGACCCCTTTAAACCTTTAAACCTTTAAACCTTTAAACCTTTAAACCTTTAA 328
|||||
DB 241 GCTAAGAGACCCCTTTAAACCTTTAAACCTTTAAACCTTTAAACCTTTAAACCTTTAA 300
OY 329 GTTGGCAATTTGTTCCGATTAAGATATATCTGAACTATGTCATCCCTGAGACT 388
|||||
DB 301 GTTGGCAATTTGTTCCGATTAAGATATATCTGAACTATGTCATCCCTGAGACT 360
OY 389 TGGGAAGACCCGACCAATGATATTAACATGATGATGATGATGATGATGATGATGATGAT 448
|||||
DB 361 TGGGAAGACCCGACCAATGATATTAACATGATGATGATGATGATGATGATGATGATGAT 420
OY 449 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
|||||
DB 421 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
OY 509 CTAGGCATATTTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 568
|||||
DB 481 CTAGGCATATTTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
OY 569 GTGATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 628
|||||
DB 541 GTGATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
OY 629 GGCTACTTAGAGCTACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
|||||
DB 601 GGCTACTTAGAGCTACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 632

RESULT 7
BE887465 628 bp mRNA EST 20-OCT-2000
LOCUS 601508111F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3909506 5',
DEFINITION mRNA sequence.
ACCESSION BE887465
VERSION BE887465.1 GI:10342782
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 628)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM9723. row: f column: 03
 High quality sequence stop: 626.
 Location/Qualifiers

FEATURES

source

1..628

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3909506"
 /clone_lib="NH_MGC_71"
 /tissue_type="Leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site: 1; Not:
 Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."
 Average insert size 2.1 kb.

BASE COUNT

203 a 107 c 151 g 167 t

ORIGIN

Query Match

Best Local Similarity 49.3%; Score 628; DB 11; Length 628;
 Matches 628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

395 GACCCAGGCGACATGATAAATGATGCTGCTGTTGTCGACATGACCAATGATGTC 454
|||||
1 GACCCAGGCGACATGATAAATGATGCTGCTGTTGTCGACATGACCAATGATGTC 60
|||||
455 TGTGAATTTGGAAGCAGATGATGTCGACAGAGTGAATTAATTTGGCTGAAAGTTCTAGGC 514
|||||
61 TGTGAATTTGGAAGCAGATGATGTCGACAGAGTGAATTAATTTGGCTGAAAGTTCTAGGC 120
|||||
515 ATATTGGCTATGATGTCGACAGAGGGAACCGACTGGAATGATGCTGATTAATGTCGAT 574
|||||
121 ATATTGGCTATGATGTCGACAGAGGGAACCGACTGGAATGATGCTGATTAATGTCGAT 180
|||||
575 GATCTGATGACACCAATTAATGATGATGTCGACAGAGGCTGAAAGCTGCTGAC 634
|||||
181 GATCTGATGACACCAATTAATGATGATGTCGACAGAGGCTGAAAGCTGCTGAC 240
|||||
635 TTGAAGCTACTGTCGCTGCTGTTAGAGGATTAAGGTTCCATGGAAGCAAGAAAT 694
|||||
241 TTGAAGCTACTGTCGCTGCTGTTAGAGGATTAAGGTTCCATGGAAGCAAGAAAT 300
|||||
695 GACTTTCGCTTATGATGATTAAGATTAAGATGATGCTGCTGATTAATTAAGC 754
|||||
301 GACTTTCGCTTATGATGATTAAGATTAAGATGATGCTGCTGATTAATTAAGC 360
|||||
755 ACTCATGACCATTTGAAGCATTAAGTACTAAGAAACGATGGAAGAAATCAGTTGC 814
|||||
361 ACTCATGACCATTTGAAGCATTAAGTACTAAGAAACGATGGAAGAAATCAGTTGC 420
|||||
815 ATGAATGACACTTTGCTGAGAGCCCTTGAAGTGTGATGCTGCTGACAGGCATT 874
|||||
421 ATGAATGACACTTTGCTGAGAGCCCTTGAAGTGTGATGCTGCTGACAGGCATT 480
|||||
875 GTGATGCTTTACACACCCCTTGATATGCTGACAGTACCAACAGAGTGATAG 934
|||||
481 GTGATGCTTTACACACCCCTTGATATGCTGACAGTACCAACAGAGTGATAG 540
|||||
935 TGTGTCATACAGAAACTATGATGATTTCTGGAATTAAGTGTATTTCTACAC 994
|||||
541 TGTGTCATACAGAAACTATGATGATTTCTGGAATTAAGTGTATTTCTACAC 600
|||||
995 TGTGTCATACAGATGATTAAGTA 1022
|||||
601 TCGGTTCATCTGATGATTAAGTA 628
|||||

```

RESULT

8

AM009649 660 bp. mRNA EST 08-MAR-2000
 LOCUS AM009649.1 NCBI CGAP Co3 Homo sapiens cDNA clone IMAGE:2504764.3
 DEFINITION Similar to SW:IPYR_BOVIN P37980 INORGANIC PYROPHOSPHATASE ;, mRNA

sequence.
 AM009649
 VERSION AM009649.1 GI:5858427
 EST.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapb@mail.nih.gov
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbip/image/image.html
 Insert Length: 418 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 457.
 Location/Qualifiers

FEATURES

source

1..660

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2504764"
 /clone_lib="NCI-CGAP-Co3"
 /sex="pooled"
 /tissue_type="colon"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site: 1; Not I; Site: 2: Eco RI; 1st strand cDNA
 was prepared from 12 pooled bulk tumor samples and primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT73 vector. Library went through one round of
 normalization."

BASE COUNT 197 a 129 c 110 g 221 t 3 others

ORIGIN

Query Match

Best Local Similarity 49.0%; Score 625; DB 10; Length 660;
 Matches 625; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

651 ACTGTTTGAAGGATTAAGTCTGATGGAACACAGAAATGATTTGGCTTAATG 710
|||||
628 ACTGTTTGAAGGATTAAGTCTGATGGAACACAGAAATGATTTGGCTTAATG 569
|||||
711 CAGATTTAAGTAAGACTTTGCCATGATTAATTAAGCACTGATGACATGGA 770
|||||
568 CAGATTTAAGTAAGACTTTGCCATGATTAATTAAGCACTGATGACATGGA 509
|||||
771 AAGCATTAGGACTAGAAAGCAATGGAAGGATGAGTGCATGATTAACATTGT 830
|||||
508 AAGCATTAGGACTAGAAAGCAATGGAAGGATGAGTGCATGATTAACATTGT 449
|||||
831 CTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAGCCATTTGATGCTTAACAC 890
|||||
448 CTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAGCCATTTGATGCTTAACAC 389
|||||
891 CACCTGTGATTTGCTGACAGTACCAAGAGCTGATAGTGTGCTTCATACACAGA 950
|||||
388 CACCTGTGATTTGCTGACAGTACCAAGAGCTGATAGTGTGCTTCATACACAGA 329
|||||
951 AAACCTAAGAGATTTCTGGAATCAAGCTGATTTCTCATCATGTTTCATCTGAT 1010
|||||

```


Db 328 AAAAATAATGAGATTTCTCTGGAATACAGCTGATATTGCTACATCTGTTTCATCTGCAT 269
OY 1011 GTATTAGAAAGTAAGTAGTAGTCTTTTCAAGCTTTTAATTTTGTAGAACATCATCTACTA 1070
Db 268 GTATTAGAAAGTAAGTAGTAGTCTTTTCAAGCTTTTAATTTTGTAGAACATCATCTACTA 209
OY 1071 AAGTAATTCGCTGCTGCTACTAATCCAAATATCTCAGAAATGTTTATCCATCTAAAGCATTTT 1130
Db 208 AAGTAATTCGCTGCTGCTACTAATCCAAATATCTCAGAAATGTTTATCCATCTAAAGCATTTT 149
OY 1131 TCATATCTCACTAGAGTAAGTATTTAGACATGCTTAATATCTCAAGAGATGTCATTTG 1190
Db 148 TCATATCTCACTAGAGTAAGTATTTAGACATGCTTAATATCTCAAGAGATGTCATTTG 89
OY 1191 GAAGTCACTTGTGATAGATGTGCAAGGAGGAGACATATTTGATGTATATGTTACCATAT 1250
Db 88 GAAGTCACTTGTGATAGATGTGCAAGGAGGAGACATATTTGATGTATATGTTACCATAT 29
OY 1251 GTTAGGAAATAAATTTATTTGCTG 1275
Db 28 GTTAGGAAATAAATTTATTTGCTG 4

RESULT 9
Bg113976 772 bp mRNA EST 08-MAY-2001
LOCUS 602674374F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4796875 5',
DEFINITION mRNA sequence.
Bg113976
ACCESSION Bg113976.1 GI:13992907
VERSION Bg113976.1 GI:13992907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 772)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.fda.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLM10682 row: C column: 20
High quality sequence stop: 767.
Location/Qualifiers

FEATURES

source

1. 772
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4796875"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gtagag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.3 kb and
normalized to R0.5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 235 a 146 c 197 g 194 t
ORIGIN

Query Match

48.5%; Score 618; DB 11; Length 772;

Best Local Similarity: 100.0%; Pred. No. 4e-308;
Matches 618; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 79 CTCGGACATATAGACGGCTTCAGACGAGAGGCGCGCCCTTCCTCGAGATA 138
Db 57 CTCGGACATATAGACGGCTTCAGACGAGAGGCGCGCCCTTCCTCGAGATA 116
OY 139 CCGAGCTCTCTCAAAAAATGAGAAAGGACATATATCTCCATTTGATATTCAT 198
Db 117 CCGAGCTCTCTCAAAAAATGAGAAAGGACATATATCTCCATTTGATATTCAT 176
OY 199 TTATGACATATAGAGATGTTGTTTACATGCTAGTACAGTACAGTACAGTAC 258
Db 177 TTATGACATATAGAGATGTTGTTTACATGCTAGTACAGTACAGTACAGTAC 236
OY 259 AATGAGATTTGCTACAAAGGACCCCTTAACCTTTAAACAGATGTGAAAGGAA 318
Db 237 AATGAGATTTGCTACAAAGGACCCCTTTAAACCTTTAAACAGATGTGAAAGGAA 296
OY 319 ACTTGCTATGTTGGCAATTTTGTCCGATATTAAGATATATCTGAACTATGTCAT 378
Db 297 ACTTGCTATGTTGGCAATTTTGTCCGATATTAAGATATATCTGAACTATGTCAT 356
OY 379 CCTCAGACTTGGGAGAACCCAGGCGACATATTAACATCTGCTGTTGTGACAA 438
Db 357 CCTCAGACTTGGGAGAACCCAGGCGACATATTAACATCTGCTGTTGTGACAA 416
OY 439 TGACCCATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 498
Db 417 TGACCCATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 476
OY 499 CGTGAAGTTCTAGCATATTTGCTATGATATGACGAGGAGGAAACGATGAAATCAT 558
Db 477 CGTGAAGTTCTAGCATATTTGCTATGATATGACGAGGAGGAAACGATGAAATCAT 536
OY 559 TGCATTTAATGTTGATGATCTGATGACGACCAATTAATGATATGATGATGATG 618
Db 537 TGCATTTAATGTTGATGATCTGATGACGACCAATTAATGATATGATGATGATG 596
OY 619 GCTGAACCTGCTACTAGAGGACGTCAGTGTGACGTTTACAAGTATTAAGTCTCGA 678
Db 597 GCTGAACCTGCTACTAGAGGACGTCAGTGTGACGTTTACAAGTATTAAGTCTCGA 656
OY 679 TGGAAACCCAGAAATGA 696
Db 657 TGGAAACCCAGAAATGA 674

RESULT 10
BF791373 955 bp mRNA EST 12-JAN-2001
LOCUS 602251332P1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4344030 5',
DEFINITION mRNA sequence.
BF791373
ACCESSION BF791373.1 GI:12096427
VERSION BF791373.1 GI:12096427
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 955)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.fda.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 932)	NIH-MGC	http://mgc.nci.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999) contact: Robert Strausberg, Ph.D.

QY	270	CTACAAAGACCCTTTAAACCTATTAAACAGAGTGTAAAAAGAAAACCTTCGTAG	329
Db	1	CTACAAAGACCCTTTAAACCTATTAAACAGAGTGTAAAAAGAAAACCTTCGTAG	60
QY	330	TTGCGATTGTCCCGTATATAAGCATATATCTGSACTATGTGTCATCCTCAGACTT	389
Db	61	TTGCGAATTTGTCCCGTATATAAGCATATATCTGSACTATGTGTCATCCTCAGACTT	120
QY	390	GGGAAGACCCGAGCGCACANTGATTAACCATPACTGGGTGTGGTGACATGACCCCAATTG	449
Db	121	GGGAAGACCCGAGCGCACANTGATTAACCATPACTGGGTGTGGTGACATGACCCCAATTG	179
QY	450	ATGTCTGTGAAATTGGAAGCAAGGTATGTGCAGAAGGTCAAAATTAATTGGCGTGAAGTTC	509
Db	180	ATGTCTGTGAAATTGGAAGCAAGGTATGTGCAGAAGGTCAAAATTAATTGGCGTGAAGTTC	239
QY	510	TAGGCATATTGGGCTATGATTTGACGGAAGGGGGAACCGCATGSAACATCTGGCATTAATG	569
Db	240	TAGGCATATTGGGCTATGATTTGACGGAAGGGGGAACCGCATGSAACATCTGGCATTAATG	299
QY	570	TGATGATATCCGATGACACCCAAATTATATGATATCATGATGTCAACGGCTGAAACCTG	629
Db	300	TGATGATATCCGATGACACCCAAATTATATGATATCATGATGTCAACGGCTGAAACCTG	359
QY	630	GCTACTTAGAAGCTACTGTGACCTGGTTAGAGGTATPAGSTTCCCTGATGSAAAACGAG	689
Db	360	GCTACTTAGAAGCTACTGTGACCTGGTTAGAGGTATPAGSTTCCCTGATGSAAAACGAG	419
QY	690	AAATGACTTTGGCTTTATGCGAATTTAAAGATTAAGACCTTCCCATGTGATATTAATA	749
Db	420	AAATGACTTTGGCTTTATGCGAATTTAAAGATTAAGACCTTCCCATGTGATATTAATA	479
QY	750	AAAGCAGTCATGACCATTTGSAAGAGATTAGTACATAGAGAAACGAAATGSAAGAGATATA	809
Db	480	AAAGCAGTCATGACCATTTGSAAGAGATTAGTACATAGAGAAACGAAATGSAAGAGATATA	539
QY	810	GTTGCATGATTAACAATTGTCTGAGAGCCCTTCAAGTGTGATCCTGATGCTGCCAGAG	869
Db	540	GTTGCATGATTAACAATTGTCTGAGAGCCCTTCAAGTGTGATCCTGATGCTGCCAGAG	599
QY	870	CCATGTGGATGCTTTACCAACACCCTGATCTGCGCGACAGTACCACAGAGCTGG	929
Db	600	CCATGTGGATGCTTTACCAACACCCTGATCTGCGCGACAGTACCACAGAGCTGG	659

Best Local Similarity 100.0%; Pred. No. 5,4e-289;
Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

39 CGTGGCGGCTGGTGGCTGTGTGGCAGCGCGCGGAGAGACTCCGGCACTATGAGGGCT 98
1 CGTGGCGGCTGGTGGCTGTGTGGCAGCGCGGAGAGACTCCGGCACTATGAGGGCT 60
99 TCAGCAGCAGGAGCGCGCGCGCGCTTCTCCCTGGAGTACGAGTCTTCCCTCAAAATG 158
61 TCAGCAGCAGGAGCGCGCGCGCGCTTCTCCCTGGAGTACGAGTCTTCCCTCAAAATG 120
159 AGAAGAGCAATATATATCTCCATTTTCATGATTTCAATTTATGAGATGATGT 218
121 AGAAGAGCAATATATATCTCCATTTTCATGATTTCAATTTATGAGATGATGT 180
219 TTCACATGTAGTGAAGTACCACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 278
181 TTCACATGTAGTGAAGTACCACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
279 ACCCTTAAACCTTATTAACAGATGTGAAAAAGAAACCTTCCCTATGTGTGCAAT 338
241 ACCCTTAAACCTTATTAACAGATGTGAAAAAGAAACCTTCCCTATGTGTGCAAT 300
339 TGTTCCTGATATAAGATATATCTGGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 398
301 TGTTCCTGATATAAGATATATCTGGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
399 CAGGAGCAATATATATCTGGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 458
361 CAGGAGCAATATATATCTGGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
459 AATTTGAGCAAGATATATCTGGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 518
421 AATTTGAGCAAGATATATCTGGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
519 TGCTATGATTTGAGCAAGATATATCTGGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 578
481 TGCTATGATTTGAGCAAGATATATCTGGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
579 CTGATGAGCAGCAATTTATATGATATATGATATGATATGATATGATATGATATGATATG 619
541 CTGATGAGCAGCAATTTATATGATATATGATATGATATGATATGATATGATATGATATGATATG 581

RESULT 17
BF966192 882 bp mRNA EST 23-JAN-2001
LOCUS 602286503F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4375297 5'
DEFINITION mRNA sequence.
ACCESSION BF966192
VERSION BF966192.1 GI:12333407
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 882)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsd-remail.nih.gov
Tissue: Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10040 row: f column: 02
High quality sequence score: 638.
Location/Qualifiers

source
1. 882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4375297"
/clone_id="NIH_MGC_95"
/tissue_type="hippocampus"
/lab="host-DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTT-3',
size-selected for average insert size 2.5 kb and
normalized to 10^5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NHRI, National
Institutes of Health). Note: this is a NIH_MGC library."

BASE COUNT 236 a 187 c 240 g 219 t
ORIGIN

Query Match 45.4%; Score 579; DB 11; Length 882;
Best Local Similarity 100.0%; Pred. No. 5,8e-288;
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

23 TTGTGAGTGGCGCGCGCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 82
8 TTGTGAGTGGCGCGCGCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 67
83 GGCATATGAGGCGCGCTGGGCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 142
68 GGCATATGAGGCGCGCTGGGCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 127
143 GTTTCCTGCAAAATGAGAAAGCAATATATATCTGCAATTTATGATATTTAT 202
128 GTTTCCTGCAAAATGAGAAAGCAATATATATCTGCAATTTATGATATTTAT 187
203 GCAGATAGAGATGT 262
188 GCAGATAGAGATGT 247
263 GAGATTTGCTACAAAGACCTTTTAAACCTTATTAACAGATATGTAAGAAAGAACTT 322
248 GAGATTTGCTACAAAGACCTTTTAAACCTTATTAACAGATATGTAAGAAAGAACTT 307
323 CGCTATGTTGCAATTTTGTCCGTATTAAGATATATCTGCAATTTATGATATTTAT 382
308 CGCTATGTTGCAATTTTGTCCGTATTAAGATATATCTGCAATTTATGATATTTAT 367
383 CAGACTTGGGAGACCCAGGAGGAGCAATATATATCTGCAATTTATGATATTTATG 442
368 CAGACTTGGGAGACCCAGGAGGAGGAGCAATATATATCTGCAATTTATGATATTTATG 427
443 CCAATTTGATGTGTGAAATTTGGAAGCAAGATATGTAAGAGTGAATTAATTTGGCTG 502
428 CCAATTTGATGTGTGAAATTTGGAAGCAAGATATGTAAGAGTGAATTAATTTGGCTG 487
503 AAAGTTCTGCAATTTTGTCTATGATTTGCAAGAGGAGAAACCGATGAAAGTCTTCC 562
488 AAAGTTCTGCAATTTTGTCTATGATTTGCAAGAGGAGAAACCGATGAAAGTCTTCC 547
563 AATTAATGTGATGATGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 601
548 AATTAATGTGATGATGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 586

RESULT 18
BG777143 711 bp mRNA EST 15-MAY-2001
LOCUS BG777143
DEFINITION 602664359F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4809591 5',
mRNA sequence.
ACCESSION BG777143
VERSION BG777143.1 GI:14047460
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. (bases 1 to 711)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov
Plate: LNCMI664 row: e column: 16
High quality sequence stop: 700.
Location/Qualifiers
1. 711
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4809591"
/clone_lib="NIH_MGC_59"
/tissue_type="mucoepidermoid carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Lung; Vector: pMDR-LIB (Clontech); Site_1: SfiI (99cgcgcgcgcgc); Site_2: SfiI (99cgcgcgcgcgc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor sequence: 5'-ATCTAGAGCGCGCGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

BASE COUNT 228 a 122 c 171 g 190 t
ORIGIN

Query Match 45.3%; Score 577; DB 11; Length 711;
Best Local Similarity 100.0%; Pred. No. 6.1e-287;
Matches 577; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 313 AGGAAACTTGGCTATGTTGGCAATTGTTCCGATTAAGATATATCTGGAACATG 372
|||||
1 AGGAAACTTGGCTATGTTGGCAATTGTTCCGATTAAGATATATCTGGAACATG 60

Db 373 TGGCAATCCCTCAGACTTGGGAAGACCCAGGACCAATGATAAATCTGCTTTGG 432
|||||
61 TGGCAATCCCTCAGACTTGGGAAGACCCAGGACCAATGATAAATCTGCTTTGG 120

Db 433 TGACATATACCAATTTGATGTGTGAATTTGAAGCAAGATATGCAAGGTGAAT 492
|||||
121 TGACATATACCAATTTGATGTGTGAATTTGAAGCAAGATATGCAAGGTGAAT 180

Db 493 AATTGGCGTGAAGTTCTAGGCAATTTGGCTATGATTTGCAAGAGGGAACCGATGGA 552
|||||
181 AATTGGCGTGAAGTTCTAGGCAATTTGGCTATGATTTGCAAGAGGGAACCGATGGA 240

Db 553 AGTCTTCCCATTTATGTGATGATCCGATGACACCAATTTATATGATATCATGATGT 612
|||||
241 AGTCTTCCCATTTATGTGATGATCCGATGACACCAATTTATATGATATCATGATGT 300

Db 613 CAAGCGGTGAAGCTGCTACTTAGAAGCTACTGTGAGACGCTTTAGAGGTATTAAGT 672
|||||
301 CAAGCGGTGAAGCTGCTACTTAGAAGCTACTGTGAGACGCTTTAGAGGTATTAAGT 360

Db 673 TCCGTATGGAAGCAAGAAATGAGTTGCGTTTAAATGCAATTTAAAGATAGAGCTT 732
|||||

Db 361 TCCGTATGGAAGCAAGAAATGAGTTGCGTTTAAATGCAATTTAAAGATAGAGCTT 420
|||||

Db 733 TCCCATGATATTTATTAAGACCTCATGACATTTGGAAGCATTTGCTAGTAAGAAAC 792
|||||

Db 421 TCCCATGATATTTATTAAGACCTCATGACATTTGGAAGCATTTGCTAGTAAGAAAC 480
|||||

Db 793 GAATGGAAGAAATCATGATTTGATGATTAACACTTGTCTGAGAGCCCTTCAAGTGA 852
|||||

Db 481 GAATGGAAGAAATCATGATTTGATGATTAACACTTGTCTGAGAGCCCTTCAAGTGA 540
|||||

Db 853 TCCGTATGCTGCCAGACCATTTGTGATGCTTTACCA 889
|||||

Db 541 TCCGTATGCTGCCAGACCATTTGTGATGCTTTACCA 577
|||||

RESULT 19
AM262869/c 575 bp mRNA EST 28-DEC-1999
LOCUS xg96e02.x1 NCI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2758490 3'
DEFINITION similar to SW:IPYR_BOVIN P37980 INORGANIC PYROPHOSPHATASE, mRNA
sequence.
ACCESSION AM262869
VERSION AM262869.1 GI:6639685
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS 1 (bases 1 to 575)
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BRGP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: www.bio.llnl.gov/bbrp/image/image.html

possible reversed clone: polyt not found
Seq primer: -400P from Gibco
High quality sequence stop: 406.
Location/Qualifiers
1. 575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2758490"
/clone_lib="NCI_CGAP_Brn53"
/tissue_type="three pooled meningiomas"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies."

BASE COUNT 175 a 109 c 99 g 192 t
ORIGIN

Query Match 45.1%; Score 575; DB 10; Length 575;
Best Local Similarity 100.0%; Pred. No. 6.5e-286;
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 701 GCGTTAATGCAAGATTTAAAGATAGGACTTGGCATGATATTTAAAGCACTGAT 760
|||||

Db 575 GCGTTAATGCAAGATTTAAAGATAGGACTTGGCATGATATTTAAAGCACTGAT 516
|||||

Db 761 GACCATTTGGAAGCAATTAAGTGAAGAAAGCAATGCAAGATGATGATGAT 820
|||||

DB 515 GACCAATGGAAAGCATTAGTACTAAGAAAAGAAATGAAAAGAAATCAGTGCATGAAAT 456
QY 821 ACAACTTGTCTGAGAGCCCTTCAAGTGTGATCTGATCTGATCTGATCTGATCTGAT 880
DB 455 ACAACTTGTCTGAGAGCCCTTCAAGTGTGATCTGATCTGATCTGATCTGATCTGAT 396
QY 881 GCTTTACCAACCCCTGTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 940
DB 395 GCTTTACCAACCCCTGTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 336
QY 941 CACCAACCAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1000
DB 335 CACCAACCAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 276
QY 1001 TCATCGATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1060
DB 275 TCATCGATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 216
QY 1061 CATCTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1120
DB 215 CATCTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 156
QY 1121 AAGCAATTTTCAATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1180
DB 155 AAGCAATTTTCAATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 96
QY 1181 TTGTCAATTTGAAAGTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1240
DB 95 TTGTCAATTTGAAAGTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 36
QY 1241 GTTACCATATGTTAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1275
DB 35 GTTACCATATGTTAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1

RESULT 20
BG702955 714 bp mRNA EST 07-MAY-2001
LOCUS 602684915f1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4817474 5',
DEFINITION mRNA sequence.
ACCESSION BG702955
VERSION BG702955.1 GI:13974812
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 714)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaab@femail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0717 row: n column: 03
High quality sequence stop: 711.
Location/Qualifiers
1. 714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4817474"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag

); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to 100.0%. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 210 a 136 c 187 g 181 t
ORIGIN

Query Match 44.9%; Score 572; DB 11; Length 714;
Best Local Similarity 100.0%; Pred. No. 2.3e-284;
Matches 572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 CTCTCTTGTCAATGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 76
DB 6 CTCTCTTGTCAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 65
QY 77 GACTCCGCGCATATGAGCGGCTTCAACGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGG 136
DB 66 GACTCCGCGCATATGAGCGGCTTCAACGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGG 125
QY 137 TACCGAGTCTTCTCTCAAAATGAGAAAGACATATATATTCATTCATGATATTTCA 196
DB 126 TACCGAGTCTTCTCTCAAAATGAGAAAGACATATATATTCATTCATGATATTTCA 185
QY 197 ATTATGCAATATGAGATGTGTTTCAATGATGATGATGATGATGATGATGATGATGAT 256
DB 186 ATTATGCAATATGAGATGTGTTTCAATGATGATGATGATGATGATGATGATGATGAT 245
QY 257 AAATGAGATATGCTCAAGAGACCCCTTAACCCATTAACCAAGATGATGATGATGATGAT 316
DB 246 AAATGAGATATGCTCAAGAGACCCCTTAACCCATTAACCAAGATGATGATGATGATGAT 305
QY 317 AAATGAGATATGCTCAAGAGACCCCTTAACCCATTAACCAAGATGATGATGATGATGAT 376
DB 306 AAATGAGATATGCTCAAGAGACCCCTTAACCCATTAACCAAGATGATGATGATGATGAT 365
QY 377 ATCCCTGACAGTGTGGAGAACCCAGGCGACATGATTAACATGATGATGATGATGATGAT 436
DB 366 ATCCCTGACAGTGTGGAGAACCCAGGCGACATGATTAACATGATGATGATGATGATGAT 425
QY 437 AATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 496
DB 426 AATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 485
QY 497 GCGGTGAAGATCTGAGCATATGATGATGATGATGATGATGATGATGATGATGATGAT 556
DB 486 GCGGTGAAGATCTGAGCATATGATGATGATGATGATGATGATGATGATGATGATGAT 545
QY 557 ATTGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 588
DB 546 ATTGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 577

RESULT 21
BG702317 769 bp mRNA EST 07-MAY-2001
LOCUS 60268363f1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4816311 5',
DEFINITION mRNA sequence.
ACCESSION BG702317
VERSION BG702317.1 GI:13973536
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 769)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLAM10714 row: m column: 16
High quality sequence stop: 764.

FEATURES

source
1. 769
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1b="IMAGE:4816311"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptPR (modified
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTTVA-3',
size-selected for average insert size 2.5 kb and
normalized to 10⁵. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 234 a 140 c 201 g 194 t
ORIGIN

Query Match 44.7%; Score 570; DB 11; Length 769;
Best Local Similarity 100.0%; Pred. No. 2.6e-283;
Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

79 CTCGGCAGTATGAGCGGCTTCAGACCGAGAGCGCGCCCTTCCTCCGTGAGTA 138
|||||
57 CTCGGCAGTATGAGCGGCTTCAGACCGAGAGCGCGCCCTTCCTCCGTGAGTA 116
|||||
139 CCGAGTCTTCTCAAAAATGAGAAAGACAATATATTCATTTTCATGATATTCAT 198
|||||
117 CCGAGTCTTCTCAAAAATGAGAAAGACAATATATTCATTTTCATGATATTCAT 176
|||||
199 TTATGAGATAGAGATGTGTTACATGATGATGATGATGATGATGATGATGAT 258
|||||
177 TTATGAGATAGAGATGTGTTACATGATGATGATGATGATGATGATGATGAT 236
|||||
259 AATGAGATGTGTAACAAGACCCCTTAACCCCTTAACCAAGATGTGTAACAAG 318
|||||
237 AATGAGATGTGTAACAAGACCCCTTAACCCCTTAACCAAGATGTGTAACAAG 296
|||||
319 ACTTGGCTATGTTGCGAATTTGTTCCGATTAAGATATATTCGAACTATGTTGC 378
|||||
297 ACTTGGCTATGTTGCGAATTTGTTCCGATTAAGATATATTCGAACTATGTTGC 356
|||||
379 CCTCAGACTTGGAGAGACCCAGGCGCATGATTAACATACGCTGTTGTGTGCA 438
|||||
357 CCTCAGACTTGGAGAGACCCAGGCGCATGATTAACATACGCTGTTGTGTGCA 416
|||||
439 TGACCCAAATGATGTGTGTAATAATGTAAGCAAGGTGTGTGCAAGAGATGTA 498
|||||
417 TGACCCAAATGATGTGTGTAATAATGTAAGCAAGGTGTGTGCAAGAGATGTA 476
|||||
499 CGTGAAGTCTAGGCTATGTTGCTATGATGAGAAAGGGAACGACATGGAAGT 558
|||||
477 CGTGAAGTCTAGGCTATGTTGCTATGATGAGAAAGGGAACGACATGGAAGT 536
|||||
559 TGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 618
|||||
537 TGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 596
|||||
619 GCTGAACCTGGCTACTTAGAAGCTACTGT 648

Db 597 GCTGAACCTGGCTACTTAGAAGCTACTGT 626

RESULT 22

LOCUS BGI80287 567 bp mRNA EST 06-FEB-2001
DEFINITION 602331122F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4432636 5',
rna sequence
ACCESSION BGI80287
VERSION BGI80287.1 GI:12686990
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 567)
AUTHORS NIH-MGC http://mgi.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLAM10189 row: k column: 05
High quality sequence stop: 565.

FEATURES

source
1. 567
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1b="IMAGE:4432636"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pCMV-Sport6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 175 a 110 c 142 g 140 t
ORIGIN

Query Match 44.5%; Score 567; DB 11; Length 567;
Best Local Similarity 100.0%; Pred. No. 8.8e-282;
Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

60 GGCAGCGCGGCGGCGGAGCTCCGGCACTATGAGCGGCTTCAGACCGAGAGCGCGCG 119
|||||
1 GGCAGCGCGGCGGCGGAGCTCCGGCACTATGAGCGGCTTCAGACCGAGAGCGCGCG 60
|||||
120 GCGCCTTCTCCCTGAGTACCGAGTCTTCTCAAAAATGAGAAAGACAATATATTC 179
|||||
61 GCGCCTTCTCCCTGAGTACCGAGTCTTCTCAAAAATGAGAAAGACAATATATTC 120
|||||
180 CATTTCATGATATTCATTAATGATTAAGATGATGATGATGATGATGATGATGAT 239
|||||
121 CATTTCATGATATTCATTAATGATTAAGATGATGATGATGATGATGATGATGAT 180
|||||
240 CAGCGTGTCTAATGCAAAAATGAGATGCTTACAAAGACCCCTTAACCTTAATAC 299
|||||
181 CAGCGTGTCTAATGCAAAAATGAGATGCTTACAAAGACCCCTTAACCTTAATAC 240
|||||
300 AAGATGGAAGAAAGAAAGAAAGCTGATGATGATGATGATGATGATGATGATGAT 359
|||||
241 AAGATGGAAGAAAGAAAGAAAGCTGATGATGATGATGATGATGATGATGATGAT 300
|||||
360 TCTGGAAGTATGTCCTCCTCAGACTTGGAAGAACCCAGGCGACATGATTAACATA 419

Db 301 TCTGAGCATATGTCCTCCACACCTTGGAGAACCCAGGCCCAATGATTAACATTA 360
Qy 420 CTGGCTGTGTGTGTAACATGACCAATGATGTGTGTAATGGAAGCAAGATATG 479
Db 361 CTGGCTGTGTGTGTAACATGACCAATGATGTGTGTAATGGAAGCAAGATATG 420
Qy 480 CAAGAGTAAATTAATGAGTGAAGTCTAGCATATGCTATGATGATGAGGAGG 539
Db 421 CAAGAGTAAATTAATGAGTGAAGTCTAGCATATGCTATGATGAGGAGG 480
Qy 540 AAACCCAGTGAAGTGAAGTGAAGTCTAGCATATGCTATGATGAGGAGG 599
Db 481 AAACCCAGTGAAGTGAAGTGAAGTCTAGCATATGCTATGATGAGGAGG 540
Qy 600 ATATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 626
Db 541 ATATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 567

RESULT 23
AM840924/c 702 bp mRNA EST 18-MAY-2000
LOCUS AM840924 R01-CN0008-290300-014-d09 CN0008 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM840924
ACCESSION AM840924
VERSION AM840924.1 GI:7934907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 702)

REFERENCE
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=4t2-R01-CN0008-290300-014-d09&t3=2000-03-29&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 649.

FEATURES

SOURCE

1. 702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CN0008"
/dev_stage="Adult"
/note="Organ: colon, normal; Vector: puc18; Site: 1: Sma1; Site: 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 203 a 147 c 121 g 231 t
ORIGIN

Query Match 44.5%; Score 567; DB 10; Length 702;
Best Local Similarity 100.0%; Pred. No. 9e-282;
Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 541 AACCCAGTGAAGTGAAGTGAAGTCTAGCATATGCTATGATGAGGAGG 600
Db 595 AACCCAGTGAAGTGAAGTGAAGTCTAGCATATGCTATGATGAGGAGG 536
Qy 601 TATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
Db 535 TATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 476
Qy 661 AAGGTATAGGTTCCGTGATGAGAAACAGAAATAGTGGTGAATGAGCAATTTAA 720
Db 475 AAGGTATAGGTTCCGTGATGAGAAACAGAAATAGTGGTGAATGAGCAATTTAA 416
Qy 721 AGATAAGCATTTGGCATGATATTTATTAAGCACTCATGATGAGCAATTTAGT 780
Db 415 AGATAAGCATTTGGCATGATATTTATTAAGCACTCATGATGAGCAATTTAGT 356
Qy 781 GACTAAGAAAGCAATGAGAAAGCAATGATGATGATGATGATGATGATGATG 840
Db 355 GACTAAGAAAGCAATGAGAAAGCAATGATGATGATGATGATGATGATGATG 296
Qy 841 CTTCAAGTGTATCTGTGATGCTGCCAGACCAATTTGATGCTTTACACACCTGTGA 900
Db 295 CTTCAAGTGTATCTGTGATGCTGCCAGACCAATTTGATGCTTTACACACCTGTGA 236
Qy 901 ATCTGCTGCACAGTACCAACAGACGTGATGATGATGATGATGATGATGATG 960
Db 235 ATCTGCTGCACAGTACCAACAGACGTGATGATGATGATGATGATGATGATG 176
Qy 961 AGATTTCCTGTGATACCAAGCTGATATGCTCATGCTGCTGATGATGATGATG 1020
Db 175 AGATTTCCTGTGATACCAAGCTGATATGCTCATGCTGCTGATGATGATGATG 116
Qy 1021 TAAAGTGTACTTTTCAACCTTTAAATTTGTAAGTCACTCATGATGATGATG 1080
Db 115 TAAAGTGTACTTTTCAACCTTTAAATTTGTAAGTCACTCATGATGATGATG 56
Qy 1081 TCGCTGTGACTATTCATATATCTCAGA 1107
Db 55 TCGCTGTGACTATTCATATATCTCAGA 29

RESULT 24
Bg179769 1076 bp mRNA EST 06-FEB-2001
LOCUS Bg179769
DEFINITION 602328856p1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4430437 5',
mRNA sequence.
ACCESSION Bg179769
VERSION Bg179769.1 GI:12686472
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1076)

REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LRAM10183 row: o column: 14
High quality sequence stop: 667.
Location/Qualifiers

FEATURES

source 1.1076
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4430437"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Prostate; Vector: PCMV-SPOrt6; Site: 1; NotCl;
Site: 2; Salt: Cloned unidirectionally; Oligo-dt primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 335 a 252 c 268 g 221 t
ORIGIN

Query Match 44.1%; Score 562; DB 11; Length 1076;
Best Local Similarity 100.0%; Pred. No. 3.6e-279;
Matches 562; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 GGGTCGTGTGACAGCGGCGGCGGAGAGCTCCGGACATGAGGGGCTTCAGACCGGAGA 111
|||
Db 1 GGGTCGTGTGACAGCGGCGGCGGAGAGCTCCGGACATGAGGGGCTTCAGACCGGAGA 60
|||
OY 112 GCGCGCGCGCGCGCTTCTCCCTGAGTACCGAGTCTTCCCTCAAAATGAGAAAGACATA 171
|||
Db 61 GCGCGCGCGCGCGCTTCTCCCTGAGTACCGAGTCTTCCCTCAAAATGAGAAAGACATA 120
|||
OY 172 TATATCTCCATTCATGATATTCATTTATGAGATTAAGATGTGTTTACATGATGAGT 231
|||
Db 121 TATATCTCCATTCATGATATTCATTTATGAGATTAAGATGTGTTTACATGATGAGT 180
|||
OY 232 TGAAGTACCAAGCTGTGCTATGCAAAATGAGATGCTACAAAGAGCCCTTTAAACC 291
|||
Db 181 TGAAGTACCAAGCTGTGCTATGCAAAATGAGATGCTACAAAGAGCCCTTTAAACC 240
|||
OY 292 TATTAACAAGATGTGAGAAAGAAAGAACTTCCCTATGTGCGAATTTGTTCCGTAA 351
|||
Db 241 TATTAACAAGATGTGAGAAAGAAAGAACTTCCCTATGTGCGAATTTGTTCCGTAA 300
|||
OY 352 AGATATATCTGGAATGATGTGCTCCCTGAGACTTGGGAGAGCCAGGACATGTA 411
|||
Db 301 AGATATATCTGGAATGATGTGCTCCCTGAGACTTGGGAGAGCCAGGACATGTA 360
|||
OY 412 TAAACATGCTGCTGTTGGTGAACAATGACCAATGATGTGTAATTTGGAACAA 471
|||
Db 361 TAAACATGCTGCTGTTGGTGAACAATGACCAATGATGTGTAATTTGGAACAA 420
|||
OY 472 GGTATGTGCAAGAGTGAATAATTTGGCGTGAAGTTCTAGGATATTTGGCTATGATGA 531
|||
Db 421 GGTATGTGCAAGAGTGAATAATTTGGCGTGAAGTTCTAGGATATTTGGCTATGATGA 480
|||
OY 532 CGAAGGGGAAAGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 591
|||
Db 481 CGAAGGGGAAAGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 540
|||
OY 592 TTATATGATATCATGATGTC 613
|||
Db 541 TTATATGATATCATGATGTC 562
|||

RESULT 25
AM873109 585 bp mRNA EST 22-MAY-2000
LOCUS AM873109/c hg22h11.x1 NCI CGAP Adrl Homo sapiens cDNA clone IMAGE:3120165 3'
DEFINITION similar to SW-IPR_BOVIN P37980 INORGANIC PYROPHOSPHATASE ;, mRNA
sequence.
ACCESSION AM873109 GI:8007162
VERSION AM873109.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 585)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@femail.nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Possible reversed clone: polyT not found
Seq primer: -400P from GIBCO
High quality sequence stop: 409.
Location/Qualifiers

FEATURES
source 1.585
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3120165"
/clone_lib="NCI_CGAP_Adrl"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: PCMV-SPOrt6; Site: 1;
NotCl; Site: 2; Salt: Cloned unidirectionally; Primer:
Oligo dt. Average insert size 1.2 kb. Library
constructed by Life Technologies."
BASE COUNT 177 a 111 c 99 g 198 t
ORIGIN

Query Match 43.7%; Score 557; DB 10; Length 585;
Best Local Similarity 100.0%; Pred. No. 1.3e-276;
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 719 AAGATTAAGCACTTGGCATGATATTTAAAGCACTCATGACCATTTGGAAGCACTTA 778
|||
Db 561 AAGATTAAGCACTTGGCATGATATTTAAAGCACTCATGACCATTTGGAAGCACTTA 502
|||
OY 779 GTGACTAAGAAAGCAATGAGAAAGAAAGCAATGATGATGATGATGATGATGATGATG 838
|||
Db 501 GTGACTAAGAAAGCAATGAGAAAGAAAGCAATGATGATGATGATGATGATGATGATG 442
|||
OY 839 CCCTTCAAGTGTGATCTGATGCTGCCAGAGCCATTTGATGCTTTACCAACACCTGT 898
|||
Db 441 CCCTTCAAGTGTGATCTGATGCTGCCAGAGCCATTTGATGCTTTACCAACACCTGT 382
|||
OY 899 GAATCTGCTGCACAGTACCAAGAGAGCTGATGATGATGATGATGATGATGATGATGATG 958
|||
Db 381 GAATCTGCTGCACAGTACCAAGAGAGCTGATGATGATGATGATGATGATGATGATGATG 322
|||
OY 959 TGAGATTTCTGTGAATCAAGCTGATATTTCTCATGCTGATGATGATGATGATGATGATG 1018
|||
Db 321 TGAGATTTCTGTGAATCAAGCTGATATTTCTCATGCTGATGATGATGATGATGATGATG 262
|||
OY 1019 AGTAAAGTACTAGCTTTTCAAGCTTTAAATTTGTAGACTCATTTACTTAAGTAAAT 1078
|||
Db 261 AGTAAAGTACTAGCTTTTCAAGCTTTAAATTTGTAGACTCATTTACTTAAGTAAAT 202
|||
OY 1079 TCTGCTGTGACTATTCATATGATGATGATGATGATGATGATGATGATGATGATGATG 1138
|||
Db 201 TCTGCTGTGACTATTCATATGATGATGATGATGATGATGATGATGATGATGATGATG 142
|||
OY 1139 CAACCTAAGATTAAGCTTTTACACATGCTTAAATATCAAGAGCTGTCTATTGGAAGTAC 1198
|||
Db 141 CAACCTAAGATTAAGCTTTTACACATGCTTAAATATCAAGAGCTGTCTATTGGAAGTAC 82
|||
OY 1199 TTGTGAATAGATGTGCAAGGGGACACATTTGATGATGATGATGATGATGATGATGATG 1258
|||

Db 81 TTGTAATAGATGTCAGGGGAGACATATTGATGTATATGTTACATATGTTAGAA 22
QY 1259 ATAAATATTGCTG 1275
Db 21 ATAAATATTGCTG 5

RESULT 26
AM469221/c 589 bp mRNA EST 24-FEB-2000
LOCUS hC79609.x1 NCI-CGAP.Gas4 Homo sapiens cDNA clone IMAGE:2898952 3'
DEFINITION Similar to SW:119YR_BOVIN P37980 INORGANIC PYROPHOSPHATASE ; mRNA
sequence.
ACCESSION AM469221 GI:7039327
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 589)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.lnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 415.
Location/Qualifiers

FEATURES
SOURCE
1..589
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2898952"
/clone_lib="NCI-CGAP.Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"
/note="Organ: Stomach; Vector: pCMV-Sport6; Site:1: Salt;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"

BASE COUNT 179 a 111 c 99 g 199 t 1 others

ORIGIN

Query Match 43.5%; Score 555; DB 10; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.4e-275;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 AGATAAGACCTTGCATGATATTATTAAGACATCATGACCATTTGGAACCATTAAG 780
Db 560 AGATAAGACCTTGCATGATATTATTAAGACATCATGACCATTTGGAACCATTAAG 501
QY 781 GACTAAGAAAGCAATGAAAGCAATGCTTCATGATATCAACTTTGCTGAGAGCCC 840
Db 500 GACTAAGAAAGCAATGAAAGCAATGCTTCATGATATCAACTTTGCTGAGAGCCC 441
QY 841 CTTCAGGTGATCTGATGCTCCAGAGCCATTTGGATGCTTTACCAACCCCTGTGA 900
Db 440 CTTCAGGTGATCTGATGCTCCAGAGCCATTTGGATGCTTTACCAACCCCTGTGA 381
QY 901 ATCTGCTGACAGTACACAGACGCTGATGCTTCATGATATCAACTTTGCTGAGAGCCC 960
Db 380 ATCTGCTGACAGTACACAGACGCTGATGCTTCATGATATCAACTTTGCTGAGAGCCC 321

QY 961 AGATTCTCGAATACACAGCTGATATTGCTACATGCTGTTACATGATGATTTAGAG 1020
Db 320 AGATTCTCGAATACACAGCTGATATTGCTACATGCTGTTACATGATGATTTAGAG 261
QY 1021 TAAAGTAGTAGCTTTTCAAGCTTTAAATTTGTAGACATCTAATCAAGTAATTC 1080
Db 260 TAAAGTAGTAGCTTTTCAAGCTTTAAATTTGTAGACATCTAATCAAGTAATTC 201
QY 1081 TCGTGTAGTAAATCCATATACATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 200 TCGTGTAGTAAATCCAT 141
QY 1141 ACTAAGATTAATCTTTTACACATGCTTAAATATCAAGACAGTGTCTATTTGAGAGTCA 1200
Db 140 ACTAAGATTAATCTTTTACACATGCTTAAATATCAAGACAGTGTCTATTTGAGAGTCA 81
QY 1201 GTGATAGATGTGCAAGGGGAGCAGATATGATGATGATGATGATGATGATGATGAT 1260
Db 80 GTGATAGATGTGCAAGGGGAGCAGATATGATGATGATGATGATGATGATGATGAT 21
QY 1261 AAAATTTATTGCTG 1275
Db 20 AAAATTTATTGCTG 6

RESULT 27
BE867111 592 bp mRNA EST 20-OCT-2000
LOCUS BE867111
DEFINITION 601442951.F1 NIH-MGC_65 Homo sapiens cDNA clone IMAGE:3846924 5',
mRNA sequence.
ACCESSION BE867111
VERSION BE867111.1 GI:10315887
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 592)
NIH-MGC http://mhc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.lnl.gov
Plate: L1AM9560 row: f column: 13
High quality sequence stop: 592.
Location/Qualifiers

FEATURES
SOURCE
1..592
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846924"
/clone_lib="NIH-MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Colon; Vector: pCMV-Sport6; Site:1: NotI;
Site:2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 174 a 117 c 154 g 147 t

ORIGIN

Query Match 43.3%; Score 552; DB 11; Length 592;
Best Local Similarity 100.0%; Pred. No. 5e-274;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 79 CTCGGCAGCTATGAGCGGCTTCAGCAGCAGGAGCGCCGCCCTTCCTCCAGAGTA 138
    |||||||
Db 41 CTCGGCAGCTATGAGCGGCTTCAGCAGCAGGAGCGCCGCCCTTCCTCCAGAGTA 100
OY 139 CCGAGCTCTCTCTCAAAAAATGAGAAAGCAATATATATCTCCATTTTCATGATTTCCAA 198
    |||||||
Db 101 CCGAGCTCTCTCTCAAAAAATGAGAAAGCAATATATATCTCCATTTTCATGATTTCCAA 160
OY 199 TTATGACAGATTAAGATGTTGTTTCACATGCTAGTGTGAAGTACCAGCGTGGTCTAATGC 258
    |||||||
Db 161 TTATGACAGATTAAGATGTTGTTTCACATGCTAGTGTGAAGTACCAGCGTGGTCTAATGC 220
OY 259 AATGAGATTTGCTCAAAAGAGACCCCTTTAAACCTATTAAACAGATGTGAAAAAGGAAA 318
    |||||||
Db 221 AATGAGATTTGCTCAAAAGAGACCCCTTTAAACCTATTAAACAGATGTGAAAAAGGAAA 280
OY 319 ACTTCGCTATGTCGGAATTTTGTCCGATTAAGAGATATATCTGGAACATATGTCGCAT 378
    |||||||
Db 281 ACTTCGCTATGTCGGAATTTTGTCCGATTAAGAGATATATCTGGAACATATGTCGCAT 340
OY 379 CCCCTCAGCTTGGGAAGACCCAGGCGCAATGATTAACATCTGCTGTTGGTGACAA 438
    |||||||
Db 341 CCCCTCAGCTTGGGAAGACCCAGGCGCAATGATTAACATCTGCTGTTGGTGACAA 400
OY 439 TGACCAATATGATGTGTGAAATTTGGAAGCAAGTATGTGCAAGAGTGAAATTAATGG 498
    |||||||
Db 401 TGACCAATATGATGTGTGAAATTTGGAAGCAAGTATGTGCAAGAGTGAAATTAATGG 460
OY 499 CGTGAAGTTCTAGGCATATTTGGCTATGATGACGAAGGGAACCGATGGAAGTAT 558
    |||||||
Db 461 CGTGAAGTTCTAGGCATATTTGGCTATGATGACGAAGGGAACCGATGGAAGTAT 520
OY 559 TGGCATTAAATGATGATGATGCTGATGCGACCAATATATATGATATCATGATGTAACG 618
    |||||||
Db 521 TGGCATTAAATGATGATGATGCTGATGCGACCAATATATATGATATCATGATGTAACG 580
OY 619 GCTGAACCTCG 630
    |||||||
Db 581 GCTGAACCTCG 592

RESULT 28
Bg721093 762 bp mRNA EST 08-MAY-2001
LOCUS 602692932F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4825206 5'
DEFINITION mRNA sequence.
ACCESSION Bg721093
VERSION Bg721093.1 GI:14000280
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 762)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: rgs@bbs.fda.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
            Toshiyuki and Piero Carninci (RIKEN)
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLN at:
            http://image.llnl.gov
            Plate: LRAM10737 row: P column: 07
            High quality sequence stop: 753.
            Location/Qualifiers
                1..762
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
FEATURES
SOURCE

```

```

/clone="IMAGE:4825206"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTATTTTATTTT-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library."
BASE COUNT 230 a 140 c 197 g 194 t 1 others
ORIGIN

Query Match 42.7%; Score 544; DB 11; Length 762;
Best Local Similarity 99.8%; Pred. No. 6.9e-270;
Matches 594; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 25 GTCACTCGCGCGCGCTGTGGGCTGTGCTCTGTGCGACGCGCGCGGACAGACTCCGG 84
    |||||||
Db 6 GTCACTCGCGCGCGCTGTGGGCTGTGCTCTGTGCGACGCGCGCGGACAGACTCCGG 65
OY 85 CACTATGAGCGGCTTCAGCAGCAGAGAGCGCGCGCGCTTCCTCGAGTACCGAGT 144
    |||||||
Db 66 CACTATGAGCGGCTTCAGCAGCAGAGAGCGCGCGCGCTTCCTCGAGTACCGAGT 125
OY 145 CTTCCTCAAAAATGAGAAAGCAATATATATCTCCATTTTCATGATATTTCAATTTATC 204
    |||||||
Db 126 CTTCCTCAAAAATGAGAAAGCAATATATATCTCCATTTTCATGATATTTCAATTTATC 185
OY 205 AGATAGAGATGTGTTCACATGCTAGTGAATACACCGCTGTCTAATGCAAAATGGA 264
    |||||||
Db 186 AGATAGAGATGTGTTCACATGCTAGTGAATACACCGCTGTCTAATGCAAAATGGA 245
OY 265 GATTCCTCAAAAGGAGACCCCTTTAAACCTATTAACAGATGTGAAAAAGAAACTTGG 324
    |||||||
Db 246 GATTCCTCAAAAGGAGACCCCTTTAAACCTATTAACAGATGTGAAAAAGAAACTTGG 305
OY 325 CTATGTTGGGAATTTGTTCCGATTAAGAGATATATCTGGAACATATGTCATCCCTCA 384
    |||||||
Db 306 CTATGTTGGGAATTTGTTCCGATTAAGAGATATATCTGGAACATATGTCATCCCTCA 365
OY 385 GACTTGGGAAGACCCAGGCGCAATGATTAACATATCTGCTGTGTGCGACAAATGACC 444
    |||||||
Db 366 GACTTGGGAAGACCCAGGCGCAATGATTAACATATCTGCTGTGTGCGACAAATGACC 425
OY 445 AATGATGTGTGTAATTTGGAAGCAAGTATGTGCAAGAGTGAATATTTGGCGTGAA 504
    |||||||
Db 426 AATGATGTGTGTAATTTGGAAGCAAGTATGTGCAAGAGTGAATATTTGGCGTGAA 485
OY 505 AGTTCTAGGCATATTTGGCTATGATGACGAAGGGAACCGAGCTGAAAGTCTTCAT 564
    |||||||
Db 486 AGTTCTAGGCATATTTGGCTATGATGACGAAGGGAACCGAGCTGAAAGTCTTCAT 545
OY 565 TAATGCGATGATTCCTGATGCGACCAATATATATGATATCATGATGCTCAACCG 619
    |||||||
Db 546 TAATGCGATGATTCCTGATGCGACCAATATATATGATATCATGATGCTCAACCG 600

RESULT 29
AA694204/c 639 bp mRNA EST 16-DEC-1997
LOCUS z142c12.s1 Soares_fetal_liver.spleen_infls.s1 Homo sapiens cDNA
DEFINITION clone IMAGE:433462 3' similar to SW:IPYR_BOVIN P37980 INORGANIC
ACCESSION AA694204
VERSION AA694204.1 GI:2695142
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


Email: c9abs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 -CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 -DNA Sequencing by: Incyte Genomics, Inc.
 -Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LHAM10671 row: 1 column: 15
 High quality sequence stop: 591.

FEATURES

SOURCE

Location/Qualifiers
 1. 593
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4792790"
 /clone_lib="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescript (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTN-3',
 size-selected for average insert size 2.3 kb and
 normalized to R0.5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 167 a 119 c 158 g 149 t
 ORIGIN

Query Match 42.0%; Score 535; DB 11; Length 593;
 Best Local Similarity 99.8%; Pred. No. 3e-265;
 Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 18 TCTCTTGTACGTGCGCGCGCGGTGCGGCTGCTGTGCGACGCGCGCGCAGG 77
 8 TCTCCTTGTACGTGCGCGCGCGGTGCGGCTGCTGTGCGACGCGCGCGCAGG 67
 78 ACTCCGCGACTGAGCGGCTGACGACGAGAGCGCGCGCGCTTCTCCCTGGAGT 137
 68 ACTCCGCGACTGAGCGGCTGACGACGAGAGCGCGCGCGCTTCTCCCTGGAGT 127
 138 ACCGAGTCTTCTCAAAAATGAGAAAGACAATATATATCTCCATTTTCATATTC 197
 128 ACCGAGTCTTCTCAAAAATGAGAAAGACAATATATATCTCCATTTTCATATTC 187
 198 TTTATGCAATGAGATGTTGTTTACATGATGTTGATGATGATGATGATGATG 257
 188 TTTATGCGATGAGATGTTGTTTACATGATGTTGATGATGATGATGATGATG 247
 258 AATGAGATGTTGATGAGATGTTGTTTACATGATGTTGATGATGATGATGATG 317
 248 AATGAGATGTTGATGAGATGTTGTTTACATGATGTTGATGATGATGATGATG 307
 318 AACTTGGATGTTGATGAGATGTTGTTTACATGATGTTGATGATGATGATG 377
 308 AACTTGGATGTTGATGAGATGTTGTTTACATGATGTTGATGATGATGATG 367
 378 TTTCTGAGCTTGGAGAGCGCGGCGACATGATGATGATGATGATGATGATG 437
 368 TTTCTGAGCTTGGAGAGCGCGGCGACATGATGATGATGATGATGATGATG 427
 438 ATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 497
 428 ATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 487
 498 GCGTGAAGTTTACGATATGTTGATGATGATGATGATGATGATGATGATG 557
 488 GCGTGAAGTTTACGATATGTTGATGATGATGATGATGATGATGATGATG 547
 558 TTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 603

Db 548 TTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 593

RESULT .33

BF033412

LOCUS

DEFINITION

mRNA sequence.

BF033412

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BF033412 914 bp mRNA EST 20-OCT-2000
 601457932F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3861727 5',
 mRNA sequence.
 BF033412 GI:10741124
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NIH-MGC http://imgc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: c9abs-remail.nih.gov
 Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LHAM598 row: 0 column: 08
 High quality sequence stop: 611.
 Location/Qualifiers
 1. 914
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3861727"
 /clone_lib="NIH_MGC_66"
 /tissue_type="adenoecarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."

FEATURES

SOURCE

BASE COUNT 277 a 172 c 225 g 240 t
 ORIGIN

Query Match 41.7%; Score 532; DB 11; Length 914;
 Best Local Similarity 99.8%; Pred. No. 1.1e-263;
 Matches 582; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

184 TCATGATATTCATATTTATGATGATGATGATGATGATGATGATGATGATG 243
 1 TCATGATATTCATATTTATGATGATGATGATGATGATGATGATGATGATG 60
 244 CTGGTCTATGCAAAAATGAGATGTTGCTACAAAGAGCCCTTAAACCTATTAACA 303
 61 CTGGTCTATGCAAAAATGAGATGTTGCTACAAAGAGCCCTTAAACCTATTAACA 120
 304 TGTGAAAAAAGGAAACTTGGCTATGTTGCGAATTTGTTCCGTTAAAGATATATCTG 363
 121 TGTGAAAAAAGGAAACTTGGCTATGTTGCGAATTTGTTCCGTTAAAGATATATCTG 180
 364 GAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 423
 181 GAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
 424 CTGTTGTGTGCAATGACCAATGATGATGATGATGATGATGATGATGATG 483
 241 CTGTTGTGTGCAATGACCAATGATGATGATGATGATGATGATGATGATG 300
 484 AGGTGAATATTTGGCGTAAAGTTCTAGCATATTTGCTATGATGATGATGATG 543

Qy 748 TAAAGCACTCATTGACCATTGGAAAGCAATTAGTGACTAAGAAAAACGAATGGAAAAAGCAAT 8075

Db 534 TAAAGCACTCATTGACCATTGGAAAGCAATTAGTGACTAAGAAAAACGAATGGAAAAAGCAAT 4755

```

FEATURES
source
    CONA Library Preparation M. Bento Soares, Ph.D.
    CONA Library Arrayed by: Greg Lennon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone distribution: NCI-CGAP clone distribution information can be
    found through the I.M.A.G.E. Consortium/MLNL at:
    www.bio.lnli.gov/dbfp/Image/Image.html
    Insert Length: 1001 Std Error: 0.00
    Seq primer: -400p from Gldco
    High quality sequence stop: 353.
    Location/Qualifiers
        1..570
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:1883873"
        /clone_lib="NCI CGAP Lu5"
        /tissue_type="carcinoid"
        /lab_host="RDH10B"
        /note="Organ: lung; Vector: pWT3D-Pac (Pharmacia) with a

```

201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai

```

FEATURES
source
location/Qualifiers
1. .690
/organism="Homo sapiens"
/cdb_xref="taxon:9606"
/clone="ADBCOF06"

```

```

/lassue_type="natural grand
/levy_stage="Adult"
/lab_host="SOLR"
/note="Vector: pbluescript sk(-); site_1: ECORI; site_2

```

BASE COUNT	224 a	121 c	153 g	190 t	2 others
ORIGIN					

Query Match	40.5%	Score 517	DB 10	Length 690
Best Local Similarity	100.0%	Pred. No.	6.2e-256	
Matches 517: Conservative	0	Mismatches	0	Indels 0
				Gaps 0

481 AAGAGTGAATTAATTGGCGTGAAGTTCTAGCATATTGGCTATGATTACGGAAGGGG 540
 49 AAGAGTGAATTAATTGGCGTGAAGTTCTAGCATATTGGCTATGATTACGGAAGGGG 60
 1 AAGAGTGAATTAATTGGCGTGAAGTTCTAGCATATTGGCTATGATTACGGAAGGGG 60

QY 541 AACGCACTGAAGATCATTTGCCATTAACTGTGATGATCCTGATGCAGCCCAATTATATGA 600
|||||
Db 61 AACCAGACTGGAAGATCATTTGCCATTAACTGTGATGATCCTGATGCAGCCCAATTATATGA .120

Dy 601 TATCAATGATGTCAAACGGCTGAACCTGGCTACTTAGAAGCTACTGTGAGCTGGTTAG 660
|||||
Dd 121 TATCATGATGTCAAACGGCTGAACCGCTACTTAGAAGCTACTGTGAGCTGGTTAG 180

QY 661 AAGTATAAGCTTCCGATGGAACACGAAAAATGATTGGCGTTAATGCAGATTAA 720
|||||
181 AAGGTATAAGCTTCCGATGGAACACGAAAAATGATTGGCGTTAATGCAGATTAA 240
Db

```

Oy 721 AGATAAGGACITTTGCCATTGATATTATTAAGAAGCCTCAGACCATTTGGAAAGCATTAGT 788
      |||||
Db 241 AGATAAGGACITTTGCCATTGATATTATTAAGAAGCCTCAGACCATTTGAAAAGCATTAGT 300

```

D_b 301 GACTAAGAAACGATGCATAAACGAATCAGTTGCATGATCAACCTTGTCTGGAGAGCCC 360

Dy. 841 CTTCAAGTGTGATCCTCGATGCTGCCAGAGCCATTGTGGATGCTTTACCAACCACCCCTGTGA 900
|||||
Dδ 361 CTTCAAGTGTGATCCTCGATGCTGCCAGAGCCATTGTGGATGCTTTACCAACCACCCCTGTGA 420

DQ 901 ATCTGCGTCACAGTACCACAAGAGCTGGATGAAGTGTTCCATCACCAGAAAACTAATG 96
|||||
|||
DB 421 ATCTGCGTCACAGTACCACAAGAGCTGGATGAAGTGTTCCATCACCAGAAAACTAATG 48

QY 961 AGATTTCCTGSAATACAAAGCTGATTTGCTACATCG 997
|||||
DB 481 AGATTTCCTGSAATACAAAGCTGATTTGCTACATCG 517

RESULT 37
BC036AA6

LOCUS	B603644.0	175 bp	1998-01-01
DEFINITION	6023325583F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4427925		
ACCESSION	M8036446		
	mRNA sequence.		

VERSION	B6030440.1	GI.12431012
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

Eukaryota; Metazoa; Chordata; Sarcopterygia; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OY	779	GTGCTAGAAAAAGCATGGAAAAAGAAATGATGGATGAAATACCACTTTGTGTGAGAC	838
Db	424	GTGACTAGAAAAAGATGGAAAAAGGATGATGTTGCATGATATCAACTTTGTGTGAGAC	365
OY	839	CCCTTCAAGTGTGATTCCTGATGCTGTCGACAGCATTGTGGATGCTTTACCCACCTGT	898
Db	364	CCCTTCAAGTGTGATTCCTGATGCTGTCGACAGCATTGTGGATGCTTTACCCACCTGT	305
OY	899	GAATCGCCTTCACAGTACCAACAGACGCTGATTAAGTGTTCATACACCGAAAACTAA	958
Db	304	GAATCGCTGCACAGTACCAACAGACGCTGATTAAGTGTTCATACACCGAAAACTAA	245
OY	959	TGAGATTTCTTGGAAATCAAGCTGGAATTTCTTACATCGTTCATCTGATGATTAAGA	1018
Db	244	TGAGATTTCTTGGAAATCAAGCTGGAATTTCTTACATCGTTCATCTGATGATTAAGA	185
OY	1019	AGTAAAGTACTAGCTTTTCAAACTTTAAATTTGTAGAACTCATCTAACTAAAGTAAAT	1078
Db	184	AGTAAAGTACTAGCTTTTCAAACTTTAAATTTGTAGAACTCATCTAACTAAAGTAAAT	125
OY	1079	TCTGCTGACTAATCCAAATTAATTAATCGAAATGTAA	1113
Db	124	TCTGCTGACTAATCCAAATTAATTAATCGAAATGTAA	90

RESULT	39
AI818185/c	
LOCUS	
DEFINITION	
AI818185	617 bp mRNA
EST	07-MAR-2000
wk2c03.x1	NCI_CGAP_Pf22 Homo sapiens cDNA clone IMAGE:2418052.3
similar to SW:IPYR_BOVIN P37980	INORGANIC PYROPHOSPHATASE ; mRNA
sequence.	

ACCESSION	A1818185
VERSION	A1818185.1
KEYWORDS	EST
SOURCE	human.

ORGANISM
Homo sapiens
Eumariyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
1 (bases 1 to 617)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ccicgap>.
WRITER
National Cancer Institute Cancer Genome Anatomy Project (CGAP).

TITLE National Cancer Institute, Cancer Outcome Inventory (1997)
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael F. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
Data Generated by: Macdonald Institute Genome Sequencing Centre

found through the I.M.A.G.E. Consortium/LNL at:
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can

www.bio.lnl.gov/bbrp/image/image.html
Insert length: 712 Std Error: 0.00
Sed primer: -40UP from Gibco

FEATURES

High quality sequence stop:	460.
Location/Qualifiers	

```
source
1. .61/
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```

/clone="IMAGE:2418052"
/clone_1ib="NCI_CGAP_Pr22"
/sex="male"

```

```

/seq=malc
/tissue_type="normal prostate"
/lab_host="DH10B"

```

/note="Organ: prostate; Vector: pV/33-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed w

a Not I -oligo(drf) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the

and cloned into the pV73 and pV73- Δ 120-125 modified pV73 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT	184 a	119 c	111 g	203 t
ORIGIN				

Query Match:	40.0%	Score 510;	DB 10;	Length 617;
Best Local Similarity:	99.8%	Pred No. 2.5e-252;		
Matches 560;	Conservative	1;	Indels	0;
			Gaps	0

QY: 715 ATTTAAGATTAGGACTTTGCCATTGATATTATTTAAAGCACTCATGACCATTGGAAAGC 774

Db 561 ATTTAAGATAGGACTTTGGCATTGTATATTATTAAAAAGCACTCAGACCCATATGGAAAGC 502

775 ATTACGACTAAGAAAACGAATGGAAAAAGCAATCAGTTGCATGAATACCAACTTTGTGA 834

Db 501 ATTACTGACCTAGCAACGAAGTGAAGAAAGATCATGTCATGAATCAACATTGTCTGA 442

QY 835 GAGCCCTTCAAGTGTGATCTCATGCTCCAGACCATTTGTGAGCTTTACACACC 894

Db 441 GAGCCCTTCAAGTGTATCCGTATGCTGGCAGAGCCATTGTGGAGCTTTCACCAACCC 382

Db 381 CTGTGATCTGGCTGCACAGTACCAACACAGTGAATAGTGTTCATCACCAGAAAA 322

QY 955 CTAATGAGATTTCCTGCAATACAGCTGATATTCGTACATCGTGTTCATCTGGATGTAT 1014

Db 321 CTAAAGAGATTTCCTCGAATACAAGCTGATATTGCTACACGCTGTTCATCTGGATGTAT 262

DB: 261 TAGAGCTAAAGCTAGTACTCTTTCAAGCTTTAAATTTGTGAGACTATCTACTAAACT 202

QY. 1075 A A A T T G C T G T G A C T A T C C A T A T A T A C T C A G A A T G T A T T C C A T C T A A G C A T T T T C A T 1134 .

Db 201 A A A T T T C G C T G T G A C T A A T C C A A T A T A C T C A G A A T G T T A T C C A T C T A A G C A T T T T T C A T 142

Db 141 ATCTCACTAAGATTAACCTTTTAGACACATGCTTAATATATCAAGACAGTTGTCATTTGGAG 82

QY 1195 TCACCTGTGAATAGATGTGCCAAGGGAGCACATATTGGATGTATGTATTACCATATGTTA 1254

Db 81 TCACITGTGAATGATGTGCCAAGGGGAGCACATATTGATGTATATGTTCACCATATGTTA 22

QY	1255	GGAAATAAAATTAATTTTCG	1275
Db	21	GGAAATAAAATTAATTTTCG	1

RESULT 40

Accession	Length	Source	EST	Date
U00001	940 bp	MRNA		13-FEB-2001
U00002		CDNA clone	IMAGE:4472715	5'

DEFINITION	mRNA sequence.	CDNA clone	INNOVATION
ACCESSION	BG254456		

```

VERSION      BG254A456.1  GI:12764272
KEYWORDS     EST.
SOURCE       human

```

	Hominidae	<i>Homo sapiens</i>
SOURCE		
ORGANISM		
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
Cnidaria; Cephalochordata; Vertebrata; Osteichthyes; Homo		

REFERENCE
AUTHORS
1 (bases 1 to 940)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
CONTACT	Robert Strausberg Ph.D.

CONTACT: ROBERT STRAUSSBERG, DIRECTOR
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP

CDNA Library Preparation by: The I.M.A.G.E. Consortium (LLNL).
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/PLNL at:

http://image.llnl.gov
Plate: LLAM10305 row: 1 column: 16
High quality sequence stop: 667.

FEATURES

Source

Location/Qualifiers

1. 940.

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4477215"

/clone_lib="NIH_MGC_91"

/tissue_type="adenocarcinoma, cell line"

/note="Organ: prostate; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dt primed."

Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC library."

BASE COUNT

275 a 183 c 232 g 250 t

ORIGIN

Query Match 39.5%; Score: 504; DB 11; Length 940;

Best Local Similarity 100.0%; Pred. No. 3.3e-249;

Matches 504; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 GGCACATGATTAACATCTGCTGTGTGTCACATGACCCCAATGATGTGTGAAA 461

Db 40 GGCACATGATTAACATCTGCTGTGTGTCACATGACCCCAATGATGTGTGAAA 99

QY 462 TTGGAGCAAGGTATGTGCAAGAGGTGAATATTTGGCGTGAAGTCTTAGCATTTGG 521

Db 100 TTGGAGCAAGGTATGTGCAAGAGGTGAATATTTGGCGTGAAGTCTTAGCATTTGG 159

QY 522 CTATGATTTGCAAGGGAAGGAGTGAAGTCAATGATGATGATGATGATGATGATG 581

Db 160 CTATGATTTGCAAGGGAAGGAGTGAAGTCAATGATGATGATGATGATGATGATG 219

QY 582 ATGCAGCAATTAATATGATATCAATGATGCAACGCGTGAACCTGCTACTTAAAG 641

Db 220 ATGCAGCAATTAATATGATATCAATGATGCAACGCGTGAACCTGCTACTTAAAG 279

QY 642 CTACTGTGAGTGTGTTAGAAAGTATTAAGTCTCTGATGCAAAACCGAAATAGTTTG 701

Db 280 CTACTGTGAGTGTGTTAGAAAGTATTAAGTCTCTGATGCAAAACCGAAATAGTTTG 339

QY 702 CGTTAATGCAAGATTTAAAGATTAAGACCTTGCATGATTAATTAAGACATGATG 761

Db 340 CGTTAATGCAAGATTTAAAGATTAAGACCTTGCATGATTAATTAAGACATGATG 399

QY 762 ACCATTGGAAGCATTAAGTACTAAGAAAGCAATGGAAGAAAGCAATGATGATGATA 821

Db 400 ACCATTGGAAGCATTAAGTACTAAGAAAGCAATGGAAGAAAGCAATGATGATGATA 459

QY 822 CACTTTGTGTGAGAGCCCTTCAAGTGTGATCTGATGCTGCGAGAGCCATTGTGATG 881

Db 460 CACTTTGTGTGAGAGCCCTTCAAGTGTGATCTGATGCTGCGAGAGCCATTGTGATG 519

QY 882 CTTTACCACCACTGTGATCTG 905

Db 520 CTTTACCACCACTGTGATCTG 543

RESULT 41

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BE887538 503 bp mRNA EST 20-OCT-2000
601508191P1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3909526 5',
mRNA sequence.
BE887538
BE887538.1 GI:10342930
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 503)

AUTHORS NIH-MGC http://img.ncbi.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs@remail.nih.gov

Tissue Procurement: ATCC

CDNA Library: Preparation: Life Technologies, Inc.

CDNA Library: Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9723 row: f column: 23

High quality sequence stop: 503.

Location/Qualifiers

1. 503

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3909526"

/clone_lib="NIH_MGC_71"

/tissue_type="leiomysarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt."

Average insert size 2.1 kb.

BASE COUNT

154 a 97 c 126 g 126 t

ORIGIN

Query Match 39.5%; Score: 503; DB 11; Length 503;

Best Local Similarity 100.0%; Pred. No. 1e-248;

Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 GCGGCGAGACTCCGCGCTATGAGCGCTTACGACCGAGAGCGCGCGCTTCT 128

Db 1 GCGGCGAGACTCCGCGCTATGAGCGCTTACGACCGAGAGCGCGCGCTTCT 60

QY 129 CCGTGAGTACCGAGTCTTCCCAAAATGAAAGCAATATATATCTCATTTGATG 188

Db 61 CCGTGAGTACCGAGTCTTCCCAAAATGAAAGCAATATATATCTCATTTGATG 120

QY 189 AATATCCAAATTTATGACATTAAGATGTTTCAATGATGATGATGATGATGATGATG 248

Db 121 AATATCCAAATTTATGACATTAAGATGTTTCAATGATGATGATGATGATGATGATG 180

QY 249 CTAATGCAAAATGAGATGCTGCTACAAAGACCTTTAAACCTTTAAACCAAGATGCA 308

Db 181 CTAATGCAAAATGAGATGCTGCTACAAAGACCTTTAAACCTTTAAACCAAGATGCA 240

QY 309 AAAAAGAAACCTTCGATGTTGGCAATTTGTTCCGATTAAGATATATCTGGAAC 368

Db 241 AAAAAGAAACCTTCGATGTTGGCAATTTGTTCCGATTAAGATATATCTGGAAC 300

QY 369 ATGGTCCATCCCTCAGACTTGGGAAGACCGAGGACAAATGATTAACATACCTGCTGT 428

Db 301 ATGGTCCATCCCTCAGACTTGGGAAGACCGAGGACAAATGATTAACATACCTGCTGT 360

QY 429 GTGTGACAAATGACCAATGATGTTGTAATGGAAGCAAGATATGCAAGAGTG 488

Db 361 GTGTGACAAATGACCAATGATGTTGTAATGGAAGCAAGATATGCAAGAGTG 420

QY 489 AATATATGCGGTGAATTTCTAGCAATTTGGCATATGATGACCAAGGGGAAACCGACT 548

Db 421 AATATATGCGGTGAATTTCTAGCAATTTGGCATATGATGACCAAGGGGAAACCGACT 480

QY 549 GGAAGCATGTCATTAATG 571

Db 481 GGAAGCATGTCATTAATG 503

RESULT 42

BE042554/c

Db 61 ATGAGATTGCTACAAAGGACCTTTAAACCTATTAACAAGATGTGAAAAAGGAAA 120
QY 320 CTTCGCTATGTTGGGAATTTGTCCTCCGTATAAGGATATATCTGGAAGTATGTCATC 379
Db 121 CTTCGCTATGTTGGGAATTTGTCCTCCGTATAAGGATATATCTGGAAGTATGTCATC 180
QY 380 CCTGAGACTTGGGAAGACCCAGGCGACAAATGATAAACAATCTGCTGTTGGTGACAT 439
Db 181 CCTGAGACTTGGGAAGACCCAGGCGACAAATGATAAACAATCTGCTGTTGGTGACAT 240
QY 440 GACCAATGATGTTGGTAATGGAATGGAAGCAAGTATGCAAGAGGTGAATATATGCG 499
Db 241 GACCAATGATGTTGGTAATGGAATGGAAGCAAGTATGCAAGAGGTGAATATATGCG 300
QY 500 GTGAAAGTTCTAGGCAATTTGGCTATGATGACGAAGGGAACCCAGTGTGAA-GTGAT 558
Db 301 GTGAAAGTTCTAGGCAATTTGGCTATGATGACGAAGGGAACCCAGTGTGAAAGTGTAT 360
QY 559 TGCCATTATGTCGATGATCTGATGACGCAATTAATGATATGATGATGCAACG 618
Db 361 TGCCATTATGTCGATGATCTGATGACGCAATTAATGATATGATGATGCAACG 420
QY 619 GCTGAACCTGCTACTTGAAGCTACTGCTGCTGTTGAGAGTATGAGTCTGCA 678
Db 421 GCTGAACCTGCTACTTGAAGCTACTGCTGCTGTTGAGAGTATGAGTCTGCA 480
QY 679 TGAAGAACCAAGAAATGATGTTGCTTAAATGACAAATTTAAAGATTAAGATGTCAT 738
Db 481 TGAAGAACCAAGAAATGATGTTGCTTAAATGACAAATTTAAAGATTAAGATGTCAT 540
QY 739 TGAATATTTAAAGCACTATGACCAATGGAACCAATTAAGTAAAGAAACCAATG 798
Db 541 TGAATATTTAAAGCACTATGACCAATGGAACCAATTAAGTAAAGAAACCAATG 600
QY 799 AAAAGGATCAGTTCGATGATAC 822
Db 601 AAAAGGATCAGTTCGATGATAC 624

RESULT 44
LOCUS B1260339 804 bp mRNA EST 17-JUL-2001
DEFINITION 602969372P1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5108993 5',
mRNA sequence.
ACCESSION B1260339
VERSION B1260339.1 GI:14818546
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 804)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1M11263 row: P column: 18
High quality sequence stop: 737.
Location/Qualifiers
1..804

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5108993"
/clone_id="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"

/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-Sport6; site_1: NotI;
site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 251 a 147 c 200 g 206 t
ORIGIN

Query Match 39.5%; Score 503; DB 11; Length 804;
Best Local Similarity 99.8%; Pred. No. 1.1e-248;
Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 75 AGAGCTCCGGCCTATGAGCGGCTTCAGCAGGAGAGCGCGCGGCTTCCTCGG 134
Db 35 AGAGCTCCGGCCTATGAGCGGCTTCAGCAGGAGAGCGCGCGGCTTCCTCGG 94
QY 135 AGTACCGAGTCTCTCTCAAAAATGAGAAAGCAATATATCTCATTTTCATATTC 194
Db 95 AGTACCGAGTCTCTCTCAAAAATGAGAAAGCAATATATCTCATTTTCATATTC 154
QY 195 CAATTATGCAATGATGATGATGTTTACATGTTGATGATGATGATGATGATG 254
Db 155 CAATTATGCAATGATGATGATGTTTACATGTTGATGATGATGATGATGATG 214
QY 255 CAAAATGAGATGCTCTACAAAGACCCCTTTAAACCTTTAAACCAAGATGTAAGAA 314
Db 215 CAAAATGAGATGCTCTACAAAGACCCCTTTAAACCTTTAAACCAAGATGTAAGAA 274
QY 315 GAAACCTGCTATGTTGGGAATTTGTTCCGCTATAAGATATATCTGAACTATG 374
Db 275 GAAACCTGCTATGTTGGGAATTTGTTCCGCTATAAGATATATCTGAACTATG 334
QY 375 CCAATCCCTGAGCTTGGGAGACCCAGGCGACAAATGATTAACATCTGCTGTTG 434
Db 335 CCAATCCCTGAGCTTGGGAGACCCAGGCGACAAATGATTAACATCTGCTGTTG 394
QY 435 ACAATGACCAATGATGATGTTGTAATTTGGAAGCAAGGATGTCGAAGGTGAATTA 494
Db 395 ACAATGACCAATGATGATGTTGTAATTTGGAAGCAAGGATGTCGAAGGTGAATTA 454
QY 495 TTGGCGTGAATTTCTAGGCAATTTGCTATGATGATGATGATGATGATGATG 554
Db 455 TTGGCGTGAATTTCTAGGCAATTTGCTATGATGATGATGATGATGATGATG 514
QY 555 TCATTGCCATTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 614
Db 515 TCATTGCCATTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 574
QY 615 AACGGCTGAACCT 628
Db 575 AACGGCTGAACCT 588

RESULT 45
LOCUS A1922084/c 624 bp mRNA EST 08-MAR-2000
DEFINITION w888b03.x1 NCI-CCAP_U1 Homo sapiens cDNA clone IMAGE:2452877 3',
similar to SW:IPR_BOVIN P37980 INORGANIC PYROPHOSPHATASE 1, mRNA
sequence.
ACCESSION A1922084
VERSION A1922084.1 GI:5658048
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 624)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 19:53:05 ; Search time 1849.16 Seconds

(without alignments)
11374.843 Million cell updates/sec

Title: US-09-415-540-2

Perfect score: 1275
Sequence: 1 CAAAGAGTTGGGCTCTCT.....CAAAATTAATATTATTTGCTG 1275

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hgo_hum:*
31: em_hgo_inv:*
32: em_htg_fod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result		Query		DB	ID	Description
No.	Score	Match	Length			
1	1275	100.0	1275	6	AR061907	AR061907 Sequence
2	1275	100.0	1275	6	AR084921	AR084921 Sequence
3	1265	99.2	1301	6	BC001022	BC001022 Homo sapi
4	1265	99.2	1303	6	AX015063	AX015063 Sequence
5	1264	99.1	1282	9	AF119665	AF119665 Homo sapi
6	1217	95.5	1239	9	AB026723	AB026723 Homo sapi
7	1185	92.9	1204	9	AF217186	AF217186 Homo sapi
8	1146	89.9	1243	9	AF154065	AF154065 Homo sapi
9	1055	82.7	1200	6	AX018061	AX018061 Sequence
10	788	61.8	856	6	AX014862	AX014862 Sequence
11	784	61.5	578	6	AF108211	AF108211 Homo sapi
12	415	32.5	578	6	AX192855	AX192855 Sequence
13	350	27.5	165756	2	AL355138	AL355138 Homo sapi
14	350	27.5	215049	2	AC067749	AC067749 Homo sapi
15	202	15.8	357	6	AX014834	AX014834 Sequence
16	158	12.4	167012	2	AC008429	AC008429 Homo sapi
17	158	12.4	213025	2	AC021159	AC021159 Homo sapi
18	130	10.2	205656	2	AC067966	AC067966 Homo sapi
19	92	7.2	195126	2	AC067746	AC067746 Homo sapi
20	92	7.2	215049	2	AC067749	AC067749 Homo sapi
21	79	6.2	155645	2	AC011012	AC011012 Homo sapi
22	74	5.8	70575	2	AC025103	AC025103 Homo sapi
23	56	4.4	1266	4	BOVIOPPP	M95283 Bovine inor
24	50	3.9	1243	10	BC010468	BC010468 Mus muscu
25	44	3.5	70575	2	AC025103	AC025103 Homo sapi
26	25	2.0	144177	9	AL513487	AL513487 Human DNA
27	22	1.7	61946	9	AF020802	AF020802 Homo sapi
28	22	1.7	184180	9	HS1042K10	AL022238 Human DNA
29	22	1.7	216425	2	AC084825	AC084825 Mus muscu
30	22	1.7	232710	2	AC073945	AC073945 Mus muscu
31	22	1.7	323000	9	HS229041	AJ229041 Homo sapi
32	22	1.7	340000	9	AP001724	AP001724 Homo sapi
33	21	1.6	958	11	CNS061A0	AL400030 T7 end of
34	21	1.6	39525	1	SCF91	AL132973 Streptomy
35	21	1.6	44548	9	AF020801	AF020801 Homo sapi
36	21	1.6	51803	9	AF176815	AF176815 Homo sapi
37	21	1.6	60159	9	AL355504	AL355504 Human DNA
38	21	1.6	99461	8	ATF4B14	AL031986 Arabidops
39	21	1.6	106007	8	ATT19K4	AL022373 Arabidops
40	21	1.6	120134	2	AC008034	AC008034 Homo sapi
41	21	1.6	123224	9	AC011414	AC011414 Homo sapi
42	21	1.6	143661	9	AC008491	AC008491 Homo sapi
43	21	1.6	146810	9	HS179115A	292540 Human DNA s
44	21	1.6	155313	2	AC068315	AC068315 Homo sapi
45	21	1.6	157007	2	AC065599	AC065599 Homo sapi

ALIGNMENTS

RESULT 1
AR061907 1275 bp DNA PAT 29-SEP-1999
LOCUS AR061907 Sequence 2 from patent US 5843665.
DEFINITION AR061907
ACCESSION AR061907
VERSION AR061907.1 GI:5989598
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1275)
AUTHORS Hawkins,P.R. and Hillman,J.L.
TITLE Human pyrophosphatase
JOURNAL Patent: US 5843665-A 2 01-DEC-1998;
FEATURES Location/Qualifiers
source 1..1275 /organism="Unknown"

BASE COUNT 394 a 235 c 294 g 351 t 1 others
ORIGIN

REFERENCE 1 (bases 1 to 1301)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk
Email: gcgaps@email.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CNA Library Preparation: Rubin Laboratory
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland,
Web Site: http://www.nisc.nih.gov/nisc_mgc/hgri.nih.gov
Contact: Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J.J., Pearson, R., Snyder, B., Stentlipop, S., Thomas, P.J., Tlionson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L., H. and Green, E.D.

FEATURES
source
1..1301
/organism="Homo sapiens"
/db_xref="LocustID:3464"
/db_xref="taxon:9606"
/clone="MGC:1221 IMAGE:3345496"
/tissue_type="Lymph, Burkitt Lymphoma"
/clone_id="NH.MGC-8"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
92..961
/codon_start=1
/product="pyrophosphatase (inorganic)"
/protein_id="AAH01022.1"
/db_xref="GI:12654395"
/translocation="MSGFSTDEBRAPPSLETRVFLKNEKQYISPHDIPYIAADKDFHMAVEPRWSNAKMEIATKDPINIKDVKRGIRYAANLEPYRGYIWNVGAITQJWEVDPGNDKHGCCGNDP IDCVEIGSKYCARKEIIGVVLGILAMIDEGDEMKVIAINVDPDANDYNDINDVYKRLPGYLEATVDMFPRYVPPGK PENEFAFNAEFKDFALIDILSTHDHMKALVYTKTKNGKISCMNTLSSPPKCPDARAIVDALPPCESACTIVPTDYDKFHHOKN"

BASE COUNT 416 a 237 c 298 g 350 t
ORIGIN

Query Match 99.28; Score 1265; DB 9; Length 1301;
Best Local Similarity 100.08; Prid. No. 0;
Matches 1265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 11 GGGGCTCTCTCTTGTACGTGCGGCCGCCGCGGCTGATGCTCTGTGCGAGCGGCGG 70
Db 14 GGGGCTCTCTCTTGTACGTGCGGCCGCCGCGGCTGATGCTCTGTGCGAGCGGCGG 73

0Y 71 CGCGAGGACCTCCGACATAGAGCGGCTTCACACCGAGAGCGCGCGCCCTTC 130
Db 74 CGCGAGGACCTCCGACATAGAGCGGCTTCACACCGAGAGCGCGCGCCCTTC 133

0Y 131 CTGGAGTCCGAGCTCTCTCAAAAATAGAAAGGACATATATATCTCCATTTGATGAT 190
Db 134 CTGGAGTCCGAGCTCTCTCAAAAATAGAAAGGACATATATATCTCCATTTGATGAT 193

OY	191	ATTCGCAATTTATGCAAGATTAAGAGATGTCTTTACATGAGTGAATGTAAGTACACAGCTGGCT	250
OY	191	ATTCGCAATTTATGCAAGATTAAGAGATGTCTTTACATGAGTGAATGTAAGTACACAGCTGGCT	250
Db	194	ATTCGCAATTTATGCAAGATTAAGAGATGTCTTTACATGAGTGAATGTAAGTACACAGCTGGCT	253
OY	251	AATGCAAAAAATGAGATTGGCTACAAAGAGACCCCTTTAAACCCATTTTAAACAAAGATGTGA	310
Db	254	AATGCAAAAAATGAGATTGGCTACAAAGAGACCCCTTTAAACCCATTTTAAACAAAGATGTGA	313
OY	311	AAAGGAAACCTGGCTATGTGGCAATTTGTCCCGATATAAAGCATATATCTGGACATAT	370
Db	314	AAAGGAAACCTGGCTATGTGGCAATTTGTCCCGATATAAAGCATATATCTGGACATAT	373
OY	371	GGTGCCATACCTCCAGACTTTGGGAAGACCCAGGCGACACATATATAAACATCTGGCTTTGT	430
Db	374	GGTGCCATACCTCCAGACTTTGGGAAGACCCAGGCGACATATATAAACATCTGGCTTTGT	433
OY	431	GGTGACATATGCCCATTTGATGTGTGTGGAATTTGGAAAGCAAGTATGTGCAGAGGTGAA	490
Db	434	GGTGACATATGCCCATTTGATGTGTGTGGAATTTGGAAAGCAAGTATGTGCAGAGGTGAA	493
OY	491	ATAATGGCGGAAGATTCTTAAGCATTTTGGCATATGATATGACCAAGGGGAAACCGACCTGG	550
Db	494	ATAATGGCGGTAAGATTCTTAAGCATTTTGGCATATGATATGACCAAGGGGAAACCGACCTGG	553
OY	551	AAAGTCATTCGCAATTAATGTGGATGATCTGTGATGACAGCCAAATTAATATGATATCAATGAT	610
Db	554	AAAGTCATTCGCAATTAATGTGGATGATCTGTGATGACAGCCAAATTAATATGATATCAATGAT	613
OY	611	GTCAAAAGCGGTGAACCTGGGTACTTATGAAGACTACTGTGACGTGGTTTGAAGAGGTATAG	670
Db	614	GTCAAAAGCGGTGAACCTGGGTACTTATGAAGACTACTGTGACGTGGTTTGAAGAGGTATAG	673
OY	671	GTTCCTGATGGAANAACCAAGAAATGAGTTCGCTTTATGCAAGATTTTAAAGTAAAGAC	730
Db	674	GTTCCTGATGGAANAACCAAGAAATGAGTTCGCTTTATGCAAGATTTTAAAGTAAAGAC	733
OY	731	TTTGCACTGATATTTATTAAGAAGACATGACCATTTGGAAGCATTTAGTACATTAAGAA	790
Db	734	TTTGCACTGATATTTATTAAGAAGACATGACCATTTGGAAGCATTTAGTACATTAAGAA	793
OY	791	ACGAATGGAANAAGGAATCAGTTGCATGAATACAACTTTGTCTGAGAGACCCCTTCAAGTGT	850
Db	794	ACGAATGGAANAAGGAATCAGTTGCATGAATACAACTTTGTCTGAGAGACCCCTTCAAGTGT	853
OY	851	GATCCTGATGCTGCCAAGACCATTTGTGTGATGCTTTACCACACCCCTGTGATCTGGCTGC	910
Db	854	GATCCTGATGCTGCCAAGACCATTTGTGTGATGCTTTACCACACCCCTGTGATCTGGCTGC	913
OY	911	ACAGTACCAACAGACGHTGATTAAGTGTGCATCCATCACCAGAAAAACTATAGATTTTCTGT	970
Db	914	ACAGTACCAACAGACGHTGATTAAGTGTGCATCCATCACCAGAAAACTATAGATTTTCTGT	973
OY	971	GGAATACAAAGCTGATTTTGTCTACATCGTGTTCATCTGAGATGTATAGAAATAAAGTACG	103
Db	974	GGAATACAAAGCTGATTTTGTCTACATCGTGTTCATCTGAGATGTATAGAAATAAAGTACG	103
OY	1031	AGCTTTTAAAGCTTTAAATTTGTAGAACTCATCTAACTAAAGTAATTTGTGCTGTGACT	1099
Db	1034	AGCTTTTAAAGCTTTAAATTTGTAGAACTCATCTAACTAAAGTAATTTGTGCTGTGACT	1099
OY	1091	AATCCAAATATCTCAGATATGTAATCATCTTAAGCATTTTTCATATCTCACTAAGATTA	115
Db	1094	AATCCAAATATCTCAGATATGTAATCATCTTAAGCATTTTTCATATCTCACTAAGATTA	115
OY	1151	CTTTAGGACATAGCTTAATATATCAAGAGAGTGTGATTTGGAGATCCTGTGAAATGAT	121
Db	1154	CTTTAGGACATAGCTTAATATATCAAGAGAGTGTGATTTGGAGATCCTGTGAAATGAT	121
OY	1211	GTGCAAGGGGAGACATATTTGAGATGTATATGTTACCATATGTTAGAAATAAATATATT	127
Db	1214	GTGCAAGGGGAGACATATTTGAGATGTATATGTTACCATATGTTAGAAATAAATATATT	127
OY	1271	TGCTG 1275	

[illegible]

Db 661 TTCCGTGATGAAACAGAAATGAGTTTGGTTTAATGAGAAATTAAGATTAAGACT 720

OY 732 TTGGCATTTGATTTATTAAGAAAGCACTGACATTTGGAAGCAATTAAGTAAAGAAA 791

Db 721 TTGGCATTTGATTTATTAAGAAAGCACTGACATTTGGAAGCAATTAAGTAAAGAAA 780

OY 792 CGAATGGAAGAAAGAAATGATTCATGATTAACAACTTTGCTGAGAGCCCTTCAAGTGTG 851

Db 781 CGAATGGAAGAAAGAAATGATTCATGATTAACAACTTTGCTGAGAGCCCTTCAAGTGTG 840

OY 852 ATCTGATGCTGCGAGAGCACTGTTGATGCTTACCAACCACTGTAATCTGCTGCA 911

Db 841 ATCTGATGCTGCGAGAGCACTGTTGATGCTTACCAACCACTGTAATCTGCTGCA 900

OY 912 CAGTACCAACAGAGCTGATTAAGTGTTCATCAACCAAGAAATTAAGATTTCTCTG 971

Db 901 CAGTACCAACAGAGCTGATTAAGTGTTCATCAACCAAGAAATTAAGATTTCTCTG 960

OY 972 GAATGCAAGCTGATTAAGTGTTCATCAACCAAGAAATTAAGATTTCTCTG 1031

Db 961 GAATGCAAGCTGATTAAGTGTTCATCAACCAAGAAATTAAGATTTCTCTG 1020

OY 1032 GCTTTCAAGCTTAATTTGTGTAAGTCACTTAAGTAAAGTAAATTTCTGCTGACTA 1091

Db 1021 GCTTTCAAGCTTAATTTGTGTAAGTCACTTAAGTAAAGTAAATTTCTGCTGACTA 1080

OY 1092 ATCCATTAATCTGATTAAGTGTTCATCAACCAAGAAATTAAGATTTCTCTG 1151

Db 1081 ATCCATTAATCTGATTAAGTGTTCATCAACCAAGAAATTAAGATTTCTCTG 1140

OY 1152 TTTTGCACATGCTTAATTTGTGTAAGTCACTTAAGTAAAGTAAATTTCTGCTGACTA 1211

Db 1141 TTTTGCACATGCTTAATTTGTGTAAGTCACTTAAGTAAAGTAAATTTCTGCTGACTA 1200

OY 1212 TCCAGGGGAGACATTTGATTAAGTGTTCATCAACCAAGAAATTAAGATTTCTCTG 1271

Db 1201 TCCAGGGGAGACATTTGATTAAGTGTTCATCAACCAAGAAATTAAGATTTCTCTG 1260

OY 1272 GCTG 1275

Db 1261 GCTG 1264

RESULT 6

AB026723 1239 bp mRNA PRI 29-SEP-1999

LOCUS AB026723

DEFINITION Homo sapiens S1D6-8061 mRNA for pyrophosphatase, complete cds.

ACCESSION AB026723.1 GI:5931601

VERSION

KEYWORDS pyrophosphatase.

SOURCE Homo sapiens tissue.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1239)

AUTHORS Saito T., Hattori A. and Miyajima N.

TITLE Putative inorganic pyrophosphatase

JOURNAL Published Only in Database (1999) In press

REFERENCE 2 (bases 1 to 1239)

AUTHORS Saito T., Hattori A. and Miyajima N.

TITLE Direct Submission

JOURNAL Submitted (26-APR-1999) to the DDBJ/EMBL/GenBank databases.

Genome Research Group; Akiyama 4-9-1, Inage-ku, Chiba 263-8555, Japan (E-mail: t.saito@nirs.go.jp, Tel: 81-43-206-3135, Fax: 81-43-251-9818)

FEATURES

Source location/Qualifiers

1..1239

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="10"

/tissue_jlb="brain"

31..900

gene

CDS

/gene="S1D6-8061"

31..900

/gene="S1D6-8061"

/codon_start=1

/product="pyrophosphatase"

/protein_id="BA04702.1"

/db_xref="GI:5931602"

/translation="MSGFSTERAPSPLEKRVFLNKGQYISPEHIDIVYDKVFEHVEPVRMSNAKMEIATKDPLEINIKODVKKGLRYVNFPPKGYIMNGVIAIPQWEHGVNDRKTKCCGNDPIDVCEISKVARELIGVVLGILMIDGELDMVIAINVDPPDPAANYNDINIVKRLKPSYLEATYDMPFRYKVPDKEPEFAFNAEKKDFDVIDIKSTHDKWKLVTIKTNGKGISCMNTLTSRPPKCPDPAARALVIALPPPCSACTVPTDVKWPHQK"

BASE COUNT 407 a 220 c 273 g 339 t

ORIGIN

Query Match 95.5% Score 1217; DB 9; Length 1239;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1217; Conservative 0;

Db 1 TGCGACGGCGGGCGGAGAGCTCCGCACTATGAGCGCTTACAGACCGAGAGCGCGCC 118

OY 59 TGGCAGCGGGCGGCGGAGAGCTCCGCACTATGAGCGCTTACAGACCGAGAGCGCGCC 118

Db 1 TGCGACGGCGGGCGGAGAGCTCCGCACTATGAGCGCTTACAGACCGAGAGCGCGCC 60

OY 119 GCGCCCTTCCCTGAGTACGAGCTTCTTCAAAAATGAGAAAGAGCAATATATCT 178

Db 61 GCGCCCTTCCCTGAGTACGAGCTTCTTCAAAAATGAGAAAGAGCAATATATCT 120

OY 179 CCATTCATATATTCATTAATTTATGAGATTAAGATTTGTTCCATGATTTGAAGTA 238

Db 121 CCATTCATATATTCATTAATTTATGAGATTAAGATTTGTTCCATGATTTGAAGTA 180

OY 239 CCAGCGTGTCTTAATGCAAAAATGAGATTTGTTCCATGATTTGAAGTA 298

Db 181 CCAGCGTGTCTTAATGCAAAAATGAGATTTGTTCCATGATTTGAAGTA 240

OY 299 CAAGATGTGAAAAAGAAAGAACTTGGTATGTCGAAATTTGTTCCATGATTTGAAGTA 358

Db 241 CAAGATGTGAAAAAGAAAGAACTTGGTATGTCGAAATTTGTTCCATGATTTGAAGTA 300

OY 359 ATCTGAACTATGCTGCTATCCCTGAGCTTTGGAGAGCCAGGCAATGATTAACAT 418

Db 301 ATCTGAACTATGCTGCTATCCCTGAGCTTTGGAGAGCCAGGCAATGATTAACAT 360

OY 419 ACTGCTGTGTGTGTCAGCAATGACCAATGATGTTGTAATTTGAAGCAAGTATCT 478

Db 361 ACTGCTGTGTGTGTCAGCAATGACCAATGATGTTGTAATTTGAAGCAAGTATCT 420

OY 479 GCAAGAGGTGAATAATTTGGCGTGAAGTTCTAGCAATTTGCTATGATTTGACGAAGG 538

Db 421 GCAAGAGGTGAATAATTTGGCGTGAAGTTCTAGCAATTTGCTATGATTTGACGAAGG 480

OY 539 GAAACCGACTGGAAGTATTTGCCATTAATGTTGATGATATCTGATGAGCAATTAAT 598

Db 481 GAAACCGACTGGAAGTATTTGCCATTAATGTTGATGATATCTGATGAGCAATTAAT 540

OY 599 GATATCAATGATGTCGCAAGGCTGAAACCTGCTACTTAGAAGCTAGTGGAGCTGTTT 658

Db 541 GATATCAATGATGTCGCAAGGCTGAAACCTGCTACTTAGAAGCTAGTGGAGCTGTTT 600

OY 659 AGAAGGTAAAGTCTGATGAGAAACAGAAATGATAGTTGGCTTAATGCAAGATTT 718

Db 601 AGAAGGTAAAGTCTGATGAGAAACAGAAATGATAGTTGGCTTAATGCAAGATTT 660

OY 719 AAAGTAAAGGCTTGGCATTTGATTTTAAAGCACTGACCAATTTGAAGCAATTA 778

Db 661 AAAGTAAAGGCTTGGCATTTGATTTTAAAGCACTGACCAATTTGAAGCAATTA 720

OY 779 GTGACTAAGAAACGAATGAGAAAGCAATGATGATGATTAACATTTGCTGAGAGC 838

Db 721 GTGACTAAGAAACGAATGAGAAAGCAATGATGATGATTAACATTTGCTGAGAGC 780

QY 839 CCCTTCAGGTGATCCTGATCTGCTCCAGAGCAGCATTTGGATGCTTTTACACCACCCCTCT 898
DB 781 CCTTCAAGTGTGATCTGATGCTCCAGAGCAGCATTTGGATGCTTTTACACCACCCCTCT 840
QY 899 GAATCTGCTGCACAGTAC 958
DB 841 GAATCTGCTGCACAGTAC 900
QY 959 TGAGATTTCTGAGATAC 1018
DB 901 TGAGATTTCTGAGATAC 960
QY 1019 AGTAAAGTAGAGCTTTTCAAGCTTTTAAATTTGTAGAACTCATCTAATAAGTAAAT 1078
DB 961 AGTAAAGTAGAGCTTTTCAAGCTTTTAAATTTGTAGAACTCATCTAATAAGTAAAT 1020
QY 1079 TCTGCTGTGACTAATTCACATATATACAGATTTATCCATCTAAGACATTTTTCATATCT 1138
DB 1021 TCTGCTGTGACTAATTCACATATATACAGATTTATCCATCTAAGACATTTTTCATATCT 1080
QY 1139 CAACATAGATCTTTTGGACATGCTTTAATATATCAAGACATTTGTCATTTGGAGTCA 1198
DB 1081 CAACATAGATCTTTTGGACATGCTTTAATATATCAAGACATTTGTCATTTGGAGTCA 1140
QY 1199 TTGTGAATAGATGTGCAAGGGGAGCAGCATATTTGATATATATACATATATAGGAA 1258
DB 1141 TTGTGAATAGATGTGCAAGGGGAGCAGCATATTTGATATATATACATATATAGGAA 1200
QY 1259 ATAAATATATTTGCTG 1275
DB 1201 ATAAATATATTTGCTG 1217

RESULT 7
AF217186 1204 bp mRNA PRI -02-DEC-2000
DEFINITION Homo sapiens inorganic pyrophosphatase 1 (PPAI) mRNA, complete cds.
ACCESSION AF217186
VERSION AF217186.1 GI:11526786
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Kanni, L., Johansson, M. and Karlsson, A.
TITLE Cloning of a human inorganic pyrophosphatase cDNA
JOURNAL Unpublished
AUTHORS Kanni, L., Johansson, M. and Karlsson, A.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999). Clinical Virology, Karolinska Institute,
Huddinge University Hospital, Stockholm S-14186, Sweden
FEATURES
source
1..1204
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1204
/gene="PPAI"
1..870
/gene="PPAI"
/note="phase-1"
/codon_start=1
/product="inorganic pyrophosphatase 1"
/protein_id="AAG36780.1"
/db_xref="GI:11526787"
/translation="MSGFSTERAPSFLEYRVFLKNEKGYISPFHDIPIYADKDF
HMYVEPRMSNAKMEIATKDLNPIKDVKGKLYLVANLEPYKGYINNYGAIPTWE
DPGHDKHTGCCGNDPIIDVCEIGSKVCAREIIGVGLIAMIIDESEDMKVIIN
VDDPDAAVNDIINDYKRLKPGYLEATVMPFRYKVPDCKPREPEFANPEFKDFAID
IKSTHDHMKALVTKTKNGKISCMNTLLSPFCDDAARAIYDALPPCESACTV
PTVDYKXPHHOKN"

BASE COUNT 403 a 210 c 258 g 333 t

ORIGIN
Query Match 92.9%; Score 1185; DB 9; Length 1204;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0;
Matches 1185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 89 ATGAGCGCTTACAGACCCAGAGAGCGCGCGCCCTTCCCTGAGTACCGAGTCTTC 148
DB 1 ATGAGCGCTTACAGACCCAGAGAGCGCGCGCCCTTCCCTGAGTACCGAGTCTTC 60
QY 149 CTCAAAATGAGAAAAGACAAATATATATCTCATTTTCATGATATTCAAATTTATGAGAT 208
DB 61 CTCAAAATGAGAAAAGACAAATATATATCTCATTTTCATGATATTCAAATTTATGAGAT 120
QY 209 AAGGATGTGTTTACATGATGTAGTGAAGTACACAGCTGGTCTATATCCAAAATATGAGATT 268
DB 121 AAGGATGTGTTTACATGATGTAGTGAAGTACACAGCTGGTCTATATCCAAAATATGAGATT 180
QY 269 GCTACAAAGGACCCCTTAAACCTTATTAACAAGATGTGAAAAAGAAAACCTTCCCTAT 328
DB 181 GCTACAAAGGACCCCTTAAACCTTATTAACAAGATGTGAAAAAGAAAACCTTCCCTAT 240
QY 329 GTTGGCAATTTGTCCCGTATTAAGGATATATCTGGAACATATGTCATCCCTGAGCT 388
DB 241 GTTGGCAATTTGTCCCGTATTAAGGATATATCTGGAACATATGTCATCCCTGAGCT 300
QY 389 TGGGAAGACCCAGGCGACAAATGATAACATCTGCTGTGTGTGTGACAAATGCCCAATT 448
DB 301 TGGGAAGACCCAGGCGACAAATGATAACATCTGCTGTGTGTGTGACAAATGCCCAATT 360
QY 449 GATGTGTGTAATTTGGAAGCAAGATATGTGCAAGAGTGAATAATTTGGCGTGAAGTT 508
DB 361 GATGTGTGTAATTTGGAAGCAAGATATGTGCAAGAGTGAATAATTTGGCGTGAAGTT 420
QY 509 CTAGCATATTTGGCTATGATTTGACGAGGGAACCCGACTGCAAGTCAATTTGCCATTAT 568
DB 421 CTAGCATATTTGGCTATGATTTGACGAGGGAACCCGACTGCAAGTCAATTTGCCATTAT 480
QY 569 GTGATGATCTCGATGACCAATTTATATGATATATGATATGATATGATATGATATGAT 628
DB 481 GTGATGATCTCGATGACCAATTTATATGATATATGATATGATATGATATGATATGAT 540
QY 629 GCTACTTATGAGATCTGCTGAGCTGTTTGAAGATATAGGTTCTGATGAGAAACCA 688
DB 541 GCTACTTATGAGATCTGCTGAGCTGTTTGAAGATATAGGTTCTGATGAGAAACCA 600
QY 689 GAAATGAGTTTGCCTTTAATGACAGAAATTTAAAGATAGGACTTTGCCATTGATATTAT 748
DB 601 GAAATGAGTTTGCCTTTAATGACAGAAATTTAAAGATAGGACTTTGCCATTGATATTAT 660
QY 749 AAAGACATCATGACCATTTGGAAGCATATGATGATTAAGAAAAGCAATGGAAGAAATC 808
DB 661 AAAGACATCATGACCATTTGGAAGCATATGATGATTAAGAAAAGCAATGGAAGAAATC 720
QY 809 AGTTCATATATACACTTTGTCTGAGAGCCCTTAAAGTGTGATCTGATGCTGCAGAA 868
DB 721 AGTTCATATATACACTTTGTCTGAGAGCCCTTAAAGTGTGATCTGATGCTGCAGAA 780
QY 869 GCCATTGTGATGCTTTTACACACACCCCTGTGAATCTGCTGCACAGTACCAACAGAGCTG 928
DB 781 GCCATTGTGATGCTTTTACACACACCCCTGTGAATCTGCTGCACAGTACCAACAGAGCTG 840
QY 929 GATTAAGTGTGATGCTTTTACACACACCCCTGTGAATCTGCTGCACAGTACCAACAGCTGAT 988
DB 841 GATTAAGTGTGATGCTTTTACACACACCCCTGTGAATCTGCTGCACAGTACCAACAGCTGAT 900
QY 989 GCTACATCTGTTTCATCTGATGATATATAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1048
DB 901 GCTACATCTGTTTCATCTGATGATATATAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 960
QY 1049 ATTTGATAGACTCATTAACATAAGTAAATTTGCTGTGACTATATCAATATATCTAGAA 1108
DB 1109 ATTTGATAGACTCATTAACATAAGTAAATTTGCTGTGACTATATCAATATATCTAGAA 1048

Db 961 ATTGTGTAAGTCACTACTAAGTAATTCGTGCTGACTAATCCATATACTCAGAA 1020
QY 1109 TGTATCATCTAAAGCATTTTTCATATCTCACTAAGATACCTTTAGACATGCTTAA 1168
Db 1021 TGTATCATCTAAAGCATTTTTCATATCTCACTAAGATACCTTTAGACATGCTTAA 1080
QY 1169 ATATCAAGAGCTGTCTGTTGGAAGTCACTTGTGAATAGATGTGCAAGGGGAGACATA 1228
Db 1081 ATATCAAGAGCTGTCTGTTGGAAGTCACTTGTGAATAGATGTGCAAGGGGAGACATA 1140
QY 1229 TTGGATGATATGTTTACCATATGTTAGAAATTAATTTATTTGTC 1273
Db 1141 TTGGATGATATGTTTACCATATGTTAGAAATTAATTTATTTGTC 1185

RESULT 8
AF154065 1243 bp mRNA PRI 09-NOV-1999
LOCUS AF154065 Homo sapiens, inorganic pyrophosphatase (pp) mRNA, complete cds.
DEFINITION AF154065
ACCESSION AF154065
VERSION AF154065.1 GI:4960207
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1243)
Fairchild, T.A. and Patejuna, G.
Cloning and expression profile of human inorganic pyrophosphatase
JOURNAL Biochim. Biophys. Acta 1447 (2-3), 133-136 (1999)
MEDLINE 20011271
REFERENCE 2 (bases 1 to 1243)
Patejuna, G.
Direct Submission
Submitted (25-MAY-1999) Molecular Cardiology Laboratory, Evanston
Northwestern Healthcare Research Institute, 2650 Ridge Avenue,
Evanston, IL 60201, USA
FEATURES
source location/Qualifiers
1..1243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q21.3"
/tissue_type="heart"
1..1243
/gene="pp"
46..915
/gene="pp"
/EC_number="3.6.1.1"
/function="catalyzes hydrolysis of pyrophosphate"
/codon_start=1
/product="inorganic pyrophosphatase"
/protein_id="A03463.1"
/db_xref="GI:4960208"
/translation="MSGSTEERAPPSLEEVFLNKGQSYSPHDIPIYADKDF
HMYVEPRMSNAKMEIATDPLNPKIDYKAGKLRVANLEPKYIWMNGAIPTOME
DEGHNDKGGCGSDNDIDYCTGSKVARGELIIGKVGITIAMDEGEMKVIAIN
YDDPDANNDINDYKRLKPGIEATVDMFRKRYVDGRENFAFAAEKDKDFRID
IIRSTDHMKALYTKTKNGISCMNTTISESPKCDPDAAAIIDALPPECSEACTV
PTVDKWFHOKN"

BASE COUNT 400 a 224 c 279 g 340 t
ORIGIN

Query Match 89.9%; Score 1146; DB 9; Length 1243;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 79 CTCGGGACATATGAGGCGCTTCAGACCGAGGAGCGCGCGCCCTTCACCTGGAGTA 138
Db 36 CTCGGGACATATGAGGCGCTTCAGACCGAGGAGCGCGCGCCCTTCACCTGGAGTA 95
QY 139 CGAGTCTTCTCTCAAAAATGAGAAAGACAATATATCTTCATTTTCATATTTCCAAAT 198

Db 96 CGAGTCTTCTCTCAAAAATGAGAAAGACAATATATATCTTCATTTTCATATTTCCAAAT 155
QY 199 TTATGACAGATTAAGAGATGTGTTTACATGTGTAGTGAAGTACACAGCTGTCTTAATGCAAA 258
Db 156 TTATGACAGATTAAGAGATGTGTTTACATGTGTAGTGAAGTACACAGCTGTCTTAATGCAAA 215
QY 259 AATGAGATTTCTCTCAAAAAGAGCCCTTTAAACCTTATTAACACATGTGAAAAAGGAAA 318
Db 216 AATGAGATTTCTCTCAAAAAGAGCCCTTTAAACCTTATTAACACATGTGAAAAAGGAAA 275
QY 319 ACTTCGATATGTTGCAATTTGTTCCGTATTAAGATATATGCAACATGTGTGTCAT 378
Db 276 ACTTCGATATGTTGCAATTTGTTCCGTATTAAGATATATGCAACATGTGTGTCAT 335
QY 439 TGACCAATGATGTGTGTGAATTTGGAAGCAAGATGTGTGCAAGAGTGAATTAATTTG 498.
Db 396 TGACCAATGATGTGTGTGAATTTGGAAGCAAGATGTGTGCAAGAGTGAATTAATTTG 455
QY 499 CGTGAAGTTCTAGCATATTTGCTATGATTTGACGAAGGGGAAACCGACATGCAAT 558
Db 456 CGTGAAGTTCTAGCATATTTGCTATGATTTGACGAAGGGGAAACCGACATGCAAT 515
QY 559 TGCCATTATGTGTGATGATCCTGATGACGCAATTAATATATATATATATATATATAT 618
Db 516 TGCCATTATGTGTGATGATCCTGATGACGCAATTAATATATATATATATATATATAT 575
QY 619 GCTGAACCTGTGCTACTTAGAAGTACTGTGACTGTTTGAAGATTAAGTTCCTGA 678
Db 576 GCTGAACCTGTGCTACTTAGAAGTACTGTGACTGTTTGAAGATTAAGTTCCTGA 635
QY 679 TGAACCAACGAAATATAGTTTGGCTTAAATGCAATTTAAAGATPAGGACTTGGCAT 738
Db 636 TGAACCAACGAAATATAGTTTGGCTTAAATGCAATTTAAAGATPAGGACTTGGCAT 695
QY 739 TGATATTTATTAAGACATGACATGACATGGAAGCAATATGATGCTAAGAAACGAATGG 798
Db 696 TGATATTTATTAAGACATGACATGACATGGAAGCAATATGATGCTAAGAAACGAATGG 755
QY 799 AAAAGGATGATGTCATGCAATTAACCTTGTGCAAGCCCTTCGAAGTGTATCTCTGA 858
Db 756 AAAAGGATGATGTCATGCAATTAACCTTGTGCAAGCCCTTCGAAGTGTATCTCTGA 815
QY 859 TGCTGCCAGAGCATTTGTGATGCTTTTACACACACCCCTGGAATCTGCTGACAGATACC 918
Db 816 TGCTGCCAGAGCATTTGTGATGCTTTTACACACACCCCTGGAATCTGCTGACAGATACC 875
QY 919 AACAGAGTGTGATGATGTTCCATCACCAGAAAACTAATGATTTCTGTGGAATACA 978
Db 876 AACAGAGTGTGATGATGTTCCATCACCAGAAAACTAATGATTTCTGTGGAATACA 935
QY 979 AGCGATATTTGCAATACGCTGTCATGATGATGATGATGATGATGATGATGATGATGATG 1038
Db 936 AGCGATATTTGCAATACGCTGTCATGATGATGATGATGATGATGATGATGATGATGATG 995
QY 1039 AAGGCTTTAATTTTGTGAGCTCATCTAATTAAGTAAATTTGCTGCTGCTAATCCAAAT 1098
Db 996 AAGGCTTTAATTTTGTGAGCTCATCTAATTAAGTAAATTTGCTGCTGCTAATCCAAAT 1055
QY 1099 ATACTGAGATGTTATTCATCTAAGCATTTTTCATATCTCAACTAAGATTAATCTTTAGC 1158
Db 1056 ATACTGAGATGTTATTCATCTAAGCATTTTTCATATCTCAACTAAGATTAATCTTTAGC 1115
QY 1159 ACATGCTTAATATCAAAAGAGAGTTGCTATTTGGAAGCACTTGGATATATATGCAAGG 1218
Db 1116 ACATGCTTAATATCAAAAGAGAGTTGCTATTTGGAAGCACTTGGATATATATGCAAGG 1175
QY 1219 GGAGCATATTTGATGATATATGTTTACCATATGTTAGAAATTAATTTATTTGCTG 1275

Db 1176 GGAGACATATGGATGATATATGTACCATATGTAGTAATAATTATTTCCTG 1232

RESULT 9
 AX018061 1200 bp DNA PAT 07-SEP-2000
 LOCUS Sequence 3 from Patent WO946374.
 DEFINITION AX018061
 ACCESSION AX018061
 VERSION AX018061.1 GI:10042512
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1200)
 AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
 Pylarsky,C.
 TITLE Human nucleic acid sequences from prostate tumour tissue
 JOURNAL Patent: WO 946374-A 3 16-SEP-1999;
 SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
 BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
 (DE); PYLARSKY CHRISTIAN (DE)
 FEATURES
 source 1. 1200
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 368 a 227 c 278 g 327 t
 ORIGIN

Query Match 82.7%; Score 1055; DB 6; Length 1200;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1055; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CCTTCCCTCCGAGTACCGAGTCTTCTCAAAAATGAGAAAGCATATATCTCCAT 182
 Db 111 CCTTCCCTCCGAGTACCGAGTCTTCTCAAAAATGAGAAAGCATATATCTCCAT 170
 QY 183 TTGATGATATTCATATTCAGATAGAGATGTTTCATCATGTTGAATACAC 242
 Db 171 TTGATGATATTCATATTCAGATAGAGATGTTTCATCATGTTGAATACAC 230
 QY 243 GCTGTTCTAATGCAAAAATGAGATTCCTACAAAAGACCTTTAAACCTATTTAAACAG 302
 Db 231 GCTGTTCTAATGCAAAAATGAGATTCCTACAAAAGACCTTTAAACCTATTTAAACAG 290
 QY 303 ATGTGAAAAAGAAAACTTCGATATGTGCGAATTTGTTCCCTATTAAGATATATCT 362
 Db 291 ATGTGAAAAAGAAAACTTCGATATGTGCGAATTTGTTCCCTATTAAGATATATCT 350
 QY 363 GGAACATATGTCATCCCTCAGACTTGGGAGAGACCCAGGCGACATATTAACATCTG 422
 Db 351 GGAACATATGTCATCCCTCAGACTTGGGAGAGACCCAGGCGACATATTAACATCTG 410
 QY 423 GCTGTTGCTGACATACCCCAATTTGATGTGTGAATTTGAAGCAAGTATGTCAA 482
 Db 411 GCTGTTGCTGACATACCCCAATTTGATGTGTGAATTTGAAGCAAGTATGTCAA 470
 QY 483 GAGGTGAATAATTTGGCGTGAAGTTCTAGGCATATTTGGCTATATGAGAGGGGAAA 542
 Db 471 GAGGTGAATAATTTGGCGTGAAGTTCTAGGCATATTTGGCTATATGAGAGGGGAAA 530
 QY 543 CCAGCTGGAAGTATTCGATTAATGTGATGATCTGATGACGCCAATTTATATGATA 602
 Db 531 CCAGCTGGAAGTATTCGATTAATGTGATGATCTGATGACGCCAATTTATATGATA 590
 QY 603 TCAATGATGTCAAACGGCTGAAACCTGCTACTAGAAAGTACTGTGACTGTTTGA 662
 Db 591 TCAATGATGTCAAACGGCTGAAACCTGCTACTAGAAAGTACTGTGACTGTTTGA 650
 QY 663 GGTATAGGTTCTGATGAGAAACAGAAATAGTTGGCTTTAATGACAAATTTAAG 722
 Db 651 GGTATAGGTTCTGATGAGAAACAGAAATAGTTGGCTTTAATGACAAATTTAAG 710

QY 723 ATAGGACTTTGCCATGATATTTAATAAGACCTCATGACCATTTGGAACCATTTAGTA 782
 Db 711 ATAGGACTTTGCCATGATATTTAATAAGACCTCATGACCATTTGGAACCATTTAGTA 770
 QY 783 CTAGAAAGCAATGAGAAAGGATTCAGTTCATGATTAACAATTTGCTGAGAGCCCT 842
 Db 771 CTAGAAAGCAATGAGAAAGGATTCAGTTCATGATTAACAATTTGCTGAGAGCCCT 830
 QY 843 TCAAGTGTATCTGATGCTGTCGAGAGCCATTTGTGATGCTTACACACCCCTGTGAAT 902
 Db 831 TCAAGTGTATCTGATGCTGTCGAGAGCCATTTGTGATGCTTACACACCCCTGTGAAT 890
 QY 903 CTGCTCCACAGTACCAACAGAGCTGATAGTGTTCATCCACAGAAACCAATTAAG 962
 Db 891 CTGCTCCACAGTACCAACAGAGCTGATAGTGTTCATCCACAGAAACCAATTAAG 950
 QY 963 ATTCTCTGAAATCAAGCTGATATTTGCTACATGCTTTCATGATGATTTAGAGTA 1022
 Db 951 ATTCTCTGAAATCAAGCTGATATTTGCTACATGCTTTCATGATGATTTAGAGTA 1010
 QY 1023 AAATGATGCTTTTCAAGCTTTAATTTGTAGAACTCATCTAAGTAAATTTCTG 1082
 Db 1011 AAATGATGCTTTTCAAGCTTTAATTTGTAGAACTCATCTAAGTAAATTTCTG 1070
 QY 1083 CTGAGCTAATCCAAATPACTCAGATGTTATCCATCTAAAGCATTTTTCATCTCAAC 1142
 Db 1071 CTGAGCTAATCCAAATPACTCAGATGTTATCCATCTAAGCATTTTTCATCTCAAC 1130
 QY 1143 TAAGTAACTTTTAGCAGATGCTTAAATATCAAG 1177
 Db 1131 TAAGTAACTTTTAGCAGATGCTTAAATATCAAG 1165

RESULT 10
 AX014862 856 bp DNA PAT 07-SEP-2000
 LOCUS Sequence 52 from Patent WO953040.
 DEFINITION AX014862
 ACCESSION AX014862
 VERSION AX014862.1 GI:10041129
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 856)
 AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
 Pylarsky,C.
 TITLE Human nucleic acid sequences from ovarian tumour tissue
 JOURNAL Patent: WO 953040-A 52 21-OCT-1999;
 SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
 BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
 (DE); PYLARSKY CHRISTIAN (DE)
 FEATURES
 source 1. 856
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 251 a 163 c 222 g 220 t
 ORIGIN

Query Match 61.8%; Score 788; DB 6; Length 856;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 788; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GGGCTCTCTCTCTGTCAGTCGGCGCGGCTGCGGCTGCTGCTGTCGACGCGG 70
 Db 17 GGGCTCTCTCTCTGTCAGTCGGCGCGGCTGCGGCTGCTGCTGTCGACGCGG 76
 QY 71 CGGAGAGACTCCGACATATAGCGGCTTACGACCGAGAGCGCGCGCCCTTCTCC 130
 Db 77 CGGAGAGACTCCGACATATAGCGGCTTACGACCGAGAGCGCGCGCCCTTCTCC 136

OY 131 CTGAGTACAGAGCTCTCTCTCAAAATGAGAAAGCAATATATATCTCCATTTCATGAT 190
 |||||
 Db 137 CTGAGTACAGAGCTCTCTCTCAAAATGAGAAAGCAATATATATCTCCATTTCATGAT 196
 OY 191 ATTCCAAATTTATGACATTAAGATGCTGTTTCATGATGTAAGTACCACGCTGTCT 250
 |||||
 Db 197 ATTCCAAATTTATGACATTAAGATGCTGTTTCATGATGTAAGTACCACGCTGTCT 256
 OY 251 AATGCAAAATGAGATTTGCTACAAAGACCCCTTTAAACCTTTAAACCAAGTGGAA 310
 |||||
 Db 257 AATGCAAAATGAGATTTGCTACAAAGACCCCTTTAAACCTTTAAACCAAGTGGAA 316
 OY 311 AAGGAAATCTGCTATGTTGCAATTTGTTCCGATTAAGATATATCTGAACTAT 370
 |||||
 Db 317 AAGGAAATCTGCTATGTTGCAATTTGTTCCGATTAAGATATATCTGAACTAT 376
 OY 371 GGTGCATCTCCATGACATTTGGGAAAGACCCAGGCAATGATAACTACTGCTGTGT 430
 |||||
 Db 377 GGTGCATCTCCATGACATTTGGGAAAGACCCAGGCAATGATAACTACTGCTGTGT 436
 OY 431 GGTGCAATGACCCATTTGATGTCTGTGAATTTGGAAGCAAGTATGTGCAAGAGTGAA 490
 |||||
 Db 437 GGTGCAATGACCCATTTGATGTGTGTGAATTTGGAAGCAAGTATGTGCAAGAGTGAA 496
 OY 491 ATAAATGGGGTGAAGTTCTAGCATATTTGGCTATGATTTGACGAAGGGGAAACGACCTGG 550
 |||||
 Db 497 ATAAATGGGGTGAAGTTCTAGCATATTTGGCTATGATTTGACGAAGGGGAAACGACCTGG 556
 OY 551 AAGTCATTTGCCATTTAATGTGATGATCTGATGACAGCAATTAATGATATCATGAT 610
 |||||
 Db 557 AAGTCATTTGCCATTTAATGTGATGATCTGATGACAGCAATTAATGATATCATGAT 616
 OY 611 GTCAAAAGGCTGAAGCTGCTACTAGAAAGTACTGTGAGTGTGTTGAAGGTAAAG 670
 |||||
 Db 617 GTCAAAAGGCTGAAGCTGCTACTAGAAAGTACTGTGAGTGTGTTGAAGGTAAAG 676
 OY 671 GTTCTGATGAAAGACAGAAATGAGTTGCTTTAATGCAAAATTAAGATAAGAC 730
 |||||
 Db 677 GTTCTGATGAAAGACAGAAATGAGTTGCTTTAATGCAAAATTAAGATAAGAC 736
 OY 731 TTTGCCATTTATTTAATAAGACATGACATTTGGAAGCATTAAGTACTAAGAA 790
 |||||
 Db 737 TTTGCCATTTATTTAATAAGACATGACATTTGGAAGCATTAAGTACTAAGAA 796
 OY 791 ACGAATGG 798
 |||||
 Db 797 ACGAATGG 804
 |||||
 RESULT 11
 AF108211 846 bp mRNA PRI 14-APR-1999
 LOCUS AF108211 Homo sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds.
 DEFINITION AF108211
 ACCESSION AF108211
 VERSION AF108211.1 GI:4583152
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 AUTHORS Rumsfeld, J., Ziegelbauer, K. and Spaltmann, F.
 TITLE Cloning, expression, affinity purification and characterization of
 polyhistidine-tagged cytosolic Saccharomyces cerevisiae and human
 inorganic pyrophosphatases for differential screening of compounds
 for antifungal activity
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 846)
 AUTHORS Rumsfeld, J., Ziegelbauer, K. and Spaltmann, F.
 TITLE Direct Submission
 JOURNAL Submitted (20-NOV-1998) Research Antinfectives 1, Bayer AG, P.O.
 Box 10179, Wuppertal 42096, Germany
 FEATURES
 Location/Qualifiers

source 1.846
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="BJAB"
 /cell_type="B-cell"
 <1..846
 /note="ipase"
 /codon_start=1
 /product="cytosolic inorganic pyrophosphatase"
 /protein_id="A02496.1"
 /db_xref="GI:4583152"
 /translation="STEERAAFLSLEKRPVPIKMGKQVSLSPHDIPYADKDFVHNV
 EEPWSNAKMEIAIKDEPLNDIKODVPLKRYANLFPYKQVIMNFGALPQWDEPGH
 NDRKTCGGNDPIDVCEISKVCAKEIIGVAVITLMIIDBETDMDKVIINDDP
 DAANYNDINVRKLPGLYELATVDWERRRYVPPGKENEFAFAEKDOPALDIKS
 TDDHKKALVTKTKTGKIGICMNTTLLSESPKCDPDAARALVDPKESACTVPTDV
 DKWHHH"
 BASE COUNT 270 a 155 c 201 g 220 t
 ORIGIN
 Query Match 61.5%; Score 784; DB 9; Length 846;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 123 CCTTCTCCCTGAGTACGAGCTCTTCTCAAAATGAGAAAGCAATATATCTCCAT 182
 |||||
 Db 23 CCTTCTCCCTGAGTACGAGCTCTTCTCAAAATGAGAAAGCAATATATCTCCAT 82
 OY 183 TTCAATGATATTCATTTATTTATGATAGATGATGTTTCCATGATGATGATGATGAT 242
 |||||
 Db 83 TTCAATGATATTCATTTATTTATGATAGATGATGTTTCCATGATGATGATGATGAT 142
 OY 243 GCTGCTAATGCAAAATGAGATGCTGCAAAAGACCCCTTTAAACCTTTAAACCAAG 302
 |||||
 Db 143 GCTGCTAATGCAAAATGAGATGCTGCAAAAGACCCCTTTAAACCTTTAAACCAAG 202
 OY 303 ATGTGAAAAAGGAAACCTGCTATGTTGCGAATTTGTTCCGATTAAGATATATCT 362
 |||||
 Db 203 ATGTGAAAAAGGAAACCTGCTATGTTGCGAATTTGTTCCGATTAAGATATATCT 262
 OY 363 GGAAGTATGTCGATCCCTCAGACTTGGAGAACCCAGGCGCAATGATTAACATACG 422
 |||||
 Db 263 GGAAGTATGTCGATCCCTCAGACTTGGAGAACCCAGGCGCAATGATTAACATACG 322
 OY 423 GCTGTTGTTGTCGATGACCAATGATGATGATGATGATGATGATGATGATGATGAT 482
 |||||
 Db 323 GCTGTTGTTGTCGATGACCAATGATGATGATGATGATGATGATGATGATGATGAT 382
 OY 483 GAGTGAATATATGCGCTGAAAGTCTAGACATATTTGCTATGATGATGATGATGATGAT 542
 |||||
 Db 383 GAGTGAATATATGCGCTGAAAGTCTAGACATATTTGCTATGATGATGATGATGATGAT 442
 OY 543 CCGACTGGAAGTCAATGCGCATTAATGATGATGATGATGATGATGATGATGATGATGAT 602
 |||||
 Db 443 CCGACTGGAAGTCAATGCGCATTAATGATGATGATGATGATGATGATGATGATGATGAT 502
 OY 603 TCAATGATGCAAAAGGCTGAAACCTGCTACTTGAAGTACATGCTGATGATGATGATGAT 662
 |||||
 Db 503 TCAATGATGCAAAAGGCTGAAACCTGCTACTTGAAGTACATGCTGATGATGATGATGAT 562
 OY 663 GGTATTAAGTCTCTGATGAGAAACAGAAATGAGATGCTGTTTAATGACAGATTTAAG 722
 |||||
 Db 563 GGTATTAAGTCTCTGATGAGAAACAGAAATGAGATGCTGTTTAATGACAGATTTAAG 622
 OY 723 ATTAAGACTTGGCATTTGATTTAATAAGACATGATGATGATGATGATGATGATGATGAT 782
 |||||
 Db 623 ATTAAGACTTGGCATTTGATTTAATAAGACATGATGATGATGATGATGATGATGATGAT 682
 OY 783 CTAGAAAGCAATGAGAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 842
 |||||
 Db 683 CTAGAAAGCAATGAGAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 742

OY	843	TCAGTGTGATCCGATGAGCTCCAGAGCCATTGTGGATGGCTTTACGACACCCCTGAT	902
Db	743	TCAGTGTGATCCGATGAGCTCCAGAGCCATTGTGGATGGCTTTACGACACCCCTGAT	802
OY	903	CTGC	906
Db	803	CTGC	806
RESULT	12		
AXI92855/c			
LOCUS	AXI92855	578 bp	DNA
DEFINITION	Sequence 422 from Patent WO0149716.		PAT
ACCESSION	AXI92855		15-AUG-2001
VERSION	AXI92855.1	GI:15210811	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 578).		
AUTHORS	Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J., Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.		
TITLE	Compounds for immunotherapy and diagnosis of colon cancer and methods for their use		
JOURNAL	Patent: WO 0149716-A 422 12-JUL-2001;		
FEATURES	CORIXA CORPORATION (us)		
SOURCE	Location/Qualifiers		
	1..578		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	155 a	139 c	99 g
ORIGIN			185 t

Query Match	32.5%	Score 415	DB 6	Length 578
Best Local Similarity	100.0%	Pred. No. 7.7e-201		
Matches 415	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0
QY	501	TGAAGTCTAGGCATATTGGCTATGATTCAGAGAGGGGGAACCCGACTGGAAGTCATG	560	
DB	415	TGAAGTCTAGGCATATTGGCTATGATTCAGAGAGGGGGAACCCGACTGGAAGTCATG	356	
QY	561	CCATTAAATGTGATGATCCCTGATGACGCCAATTATATGATATCAATGATGTCMAACGGC	620	
DB	355	CCATTAAATGTGATGATCCCTGATGACGCCAATTATATGATATCAATGATGTCMAACGGC	296	
QY	621	TGAAACCTGGCTACTTAGAAGCTACTGTGTGACTGGTTTGAAGGATTAAGTTCCTGATG	680	
DB	295	TGAAACCTGGCTACTTAGAAGCTACTGTGTGACTGGTTTGAAGGATTAAGTTCCTGATG	236	
QY	681	GAAACCCGAAATAGATTTGGCTTTAATGACGAATTTAAGATGAGCACTTGGCATG	740	
DB	235	GAAACCCGAAATAGATTTGGCTTTAATGACGAATTTAAGATGAGCACTTGGCATG	176	
QY	741	ATATTATTTAAAGCACTCATGACCAATTTGAGAAAGCATTAGTACTAGTAAGAAAGCAATGGA	800	
DB	175	ATATTATTTAAAGCACTCATGACCAATTTGAGAAAGCATTAGTACTAGTAAGAAAGCAATGGA	116	
QY	801	AAGGATCAGTTGCATGAATACAACTTTCTGAGAGCCCTTCAAGTGTGATCCGTGATG	860	
DB	115	AAGGATCAGTTGCATGAATACAACTTTCTGAGAGCCCTTCAAGTGTGATCCGTGATG	56	
QY	861	CTGCCAGAGCCATTTGTGATGCTTTACCAACCACTGTGAATCTGGCTCCACAGT	915	
DB	55	CTGCCAGAGCCATTTGTGATGCTTTACCAACCACTGTGAATCTGGCTCCACAGT	1	

ACCESSION	AL355138
VERSION	AL355138.18
KEYWORDS	HTGS_15131226
SOURCE	HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP. human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 165756) Lovell, J.
AUTHORS	Direct Submission
JOURNAL	Submitted (01-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk requests: clonerequests@sanger.ac.uk
COMMENT	On Aug 9, 2001 this sequence version replaced gi:15020403.

```

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humgen@yesanger.ac.uk
----- Project Information -----
Center project name: BA203J14
----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 160553 bases at least Q40
Consensus quality: 161175 bases at least Q30
Consensus quality: 161440 bases at least Q20
Insert size: 165756; sum-of-confids
Insert size: 158365; 9.5% error; agarose-fp
Quality coverage: 6.40x in Q20 bases; sum-of-confids
Coverage: 6.85x in Q20 bases; agarose-fp

```

```

*-----*
* NOTE: This is a 'working draft' sequence.      ✓
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

FEATURES
SOURCE

```
misc-feature
1. 165756
/noise="assembly_fragment:02723
clone_end:sp6
vector_side:right"
BASE COUNT      46322 a      38804 c      37406 g      43293 t
ORIGIN
1 others
```

	Query Match	27.5%	Score 350	DB 2	Length 165756	
	Best Local Similarity	100.0%	Pred. No. 1.1e-167			
	Matches 350	Conservative	0	Mismatches 0	Indels 0	Gaps 0
QY	936	GTGGATTAAGTGGTTCATCCACCAGAAAACTAATGAGATTTCTCTGGAATACAGCTGAT				985
Db	22335	GTGGATTAAGTGGTTCATCCACCAGAAAACTAATGAGATTTCTCTGGAATACAGCTGAT				22276
QY	986	ATTGCTACATCTGTTCATCTGTGATGTATTAGAAAGTAAAGTAAAGCTTTTCAAAGCTT				1045
Db	22275	ATTGCTACATCTGTTCATCTGTGATGTATTAGAAAGTAAAGTAAAGCTTTTCAAAGCTT				22216
QY	1046	TAAATTTGAGAACATCACTAAGTAAATTTGCTGCTAGCTAATCAATATACTCA				1105
Db	22215	TAAATTTGAGAACATCACTAAGTAAATTTGCTGCTAGCTAATCAATATACTCA				22156
QY	1106	GATGTTATCATCTTAAGCATTTTTTCATATCTCAACTAAGATTAACCTTTAGCACATGCT				1165
Db	22155	GATGTTATCATCTTAAGCATTTTTTCATATCTCAACTAAGATTAACCTTTAGCACATGCT				22096
QY	1166	TAAATATCAAGACATTTGTCATTTTGGAAGTCACTGTGTAATGATGTGCAAGGGAGACAC				1225


```
misc_feature 10542..11663
/note="assembly_name:Contig73"
misc_feature 11764..12926
/note="assembly_name:Contig75"
misc_feature 13027..14168
/note="assembly_name:Contig112"
misc_feature 14269..15574
/note="assembly_name:Contig113"
misc_feature 15675..17038
/note="assembly_name:Contig116"
misc_feature 17139..18401
/note="assembly_name:Contig118"
misc_feature 18502..19788
/note="assembly_name:Contig121"
misc_feature 19889..21245
/note="assembly_name:Contig124"
misc_feature 21346..23067
/note="assembly_name:Contig127"
misc_feature 23168..25167
/note="assembly_name:Contig134"
clone_end:SP6"
misc_feature 25268..28266
/note="assembly_name:Contig135"
misc_feature 28367..30737
/note="assembly_name:Contig136"
misc_feature 30838..35830
/note="assembly_name:Contig137"
misc_feature 35931..43873
/note="assembly_name:Contig138"
misc_feature 43974..54313
/note="assembly_name:Contig139"
misc_feature 54414..62524
/note="assembly_name:Contig140"
misc_feature 62625..75784
/note="assembly_name:Contig141"
misc_feature 75885..92445
/note="assembly_name:Contig142"
misc_feature 92546..109803
/note="assembly_name:Contig143"
misc_feature 109904..128494
/note="assembly_name:Contig144"
misc_feature 128595..146739
/note="assembly_name:Contig145"
misc_feature 146840..167230
/note="assembly_name:Contig146"
misc_feature 167331..187365
/note="assembly_name:Contig147"
misc_feature 187466..215049
/note="assembly_name:Contig148"
clone_end:T7"

BASE COUNT 57394 a 47838 c 47126 g 59466 t 3225 others
ORIGIN
```

```
Query Match 27.5%; Score 350; DB 2; Length 215049;
Best Local Similarity 100.0%; Pred. No. 1.1e-167;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 926 GTGATAGTGGTTCATCACCAGAAAACATATGAGATTTCTCGATATCAAGCTGAT 985
|||||
DB 113781 GTGATAGTGGTTCATCACCAGAAAACATATGAGATTTCTCGATATCAAGCTGAT 113840
|||||
QY 986 ATTGCTACATCGTTCATCTGAGATGATTAAGTAAAGTAGTAGCTTTTCAAGCTT 1045
|||||
DB 113841 ATTGCTACATCGTTCATCTGAGATGATTAAGTAAAGTAGTAGCTTTTCAAGCTT 113900
|||||
QY 1046 TAAATTTGAGACATCACTACTAAGTAATTCGCTGAGTAAATCAATATATACCTA 1105
|||||
DB 113901 TAAATTTGAGACATCACTACTAAGTAATTCGCTGAGTAAATCAATATATACCTA 113960
|||||
QY 1106 GAATGTTATCCATCTAAGCATTTTTCATATCTCACTAAGTAACTTTTACACATGCT 1165
|||||
DB 113961 GAATGTTATCCATCTAAGCATTTTTCATATCTCACTAAGTAACTTTTACACATGCT 114020
|||||
```

```
QY 1166 TAAATATCAACACAGTTCATTTGGAACTCTGTGAATAGATGCAAGGAGCAC 1225
|||||
DB 114021 TAAATATCAACACAGTTCATTTGGAACTCTGTGAATAGATGCAAGGAGCAC 114080
|||||
QY 1226 AATTGGAGTGTATGTTTACCATATGCTTAGGAAATTAATTTTTCGCG 1275
|||||
DB 114081 AATTGGAGTGTATGTTTACCATATGCTTAGGAAATTAATTTTTCGCG 114130
|||||
```

RESULT 15

```
AX014834 357 bp DNA PAT 07-SEP-2000
LOCUS AX014834 Sequence 23 from Patent WO953040.
ACCESSION AX014834
VERSION AX014834.1 GI:10041101
KEYWORDS
```

SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

AUTHORS

Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and

Pilarsky,C.

Human nucleic acid sequences from ovarian tumour tissue

Patent: WO 953040-A 23 21-OCT-1999;

SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN

BERND (DE); ROSENTHAL ANDRE (DE); METZGER GRS FUER GENOMFORSCHUN

(DE); PILARSKI CHRISTIAN (DE)

FEATURES

source

1..357

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 74 a 89 c 96 g 98 t

ORIGIN

```
Query Match 15.8%; Score 202; DB 6; Length 357;
Best Local Similarity 100.0%; Pred. No. 6.2e-92;
Matches 202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 11 GGGGCTCTCTCTTGTGACGTGGCGCGCGTGGGGCTGTGGACGCGCGG 70
|||||
DB 17 GGGGCTCTCTCTTGTGACGTGGCGCGCGTGGGGCTGTGGACGCGCGG 76
|||||
QY 71 CGGAGGACTCCGGGACATGAGCGGCTTCAGACGAGAGGCGCGCGCTTCTCC 130
|||||
DB 77 CGGAGGACTCCGGGACATGAGCGGCTTCAGACGAGAGGCGCGCGCTTCTCC 136
|||||
QY 131 CTGGAGTACGAGTCTTCTCAAAAATGAGAAAGCAATATATCTCATTTATGAT 190
|||||
DB 137 CTGGAGTACGAGTCTTCTCAAAAATGAGAAAGCAATATATCTCATTTATGAT 196
|||||
QY 191 ATTCCAAATTATGACAGATAAG 212
|||||
DB 197 ATTCCAAATTATGACAGATAAG 218
|||||
```

RESULT 16

```
AC008429 167012 bp DNA HTG 20-APR-2001
LOCUS AC008429 Homo sapiens chromosome 5 clone CTC-308R20, WORKING DRAFT SEQUENCE,
DEFINITION 9 unordered pieces.
```

ACCESSION

AC008429

VERSION

AC008429.4 GI:13699351

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

AUTHORS

DOE Joint Genome Institute.

Sequencing of Human Chromosome 5

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1. 1384: contig of 1384 bp in length
* 1385 1484: gap of unknown length
* 1485 3382: contig of 1898 bp in length
* 3383 3482: gap of unknown length
* 3483 5013: contig of 1531 bp in length
* 5014 5113: gap of unknown length
* 5114 6322: contig of 1209 bp in length
* 6323 6422: gap of unknown length
* 6423 7736: contig of 1314 bp in length
* 7737 7836: gap of unknown length
* 7837 9198: contig of 1362 bp in length
* 9199 9298: gap of unknown length
* 9299 11720: contig of 2422 bp in length
* 11721 11820: gap of unknown length
* 11821 13498: contig of 1678 bp in length
* 13499 13598: gap of unknown length
* 13599 15122: contig of 1524 bp in length
* 15123 15222: gap of unknown length
* 15223 17333: contig of 2110 bp in length
* 17333 17433: gap of unknown length
* 17433 19823: contig of 2391 bp in length
* 19824 19924: gap of unknown length
* 19924 21811: contig of 1888 bp in length
* 21812 21911: gap of unknown length
* 21912 24564: contig of 2653 bp in length
* 24565 24664: gap of unknown length
* 24665 27708: contig of 3044 bp in length
* 27709 30178: gap of unknown length
* 30178 30276: contig of 2368 bp in length
* 30277 33515: gap of unknown length
* 33515 35151: contig of 3239 bp in length
* 35151 35615: gap of unknown length

```

```
* 33616 37205: contig of 3590 bp in length
* 37306 37305: gap of unknown length
* 39846 39846: contig of 2541 bp in length
* 39947 39946: gap of unknown length
* 42385 42384: contig of 2438 bp in length
* 42485 42484: gap of unknown length
* 46361 46360: contig of 3776 bp in length
* 49107 49107: contig of 2747 bp in length
* 49108 49207: gap of unknown length
* 49208 54501: contig of 5294 bp in length
* 54502 54601: gap of unknown length
* 54602 59315: contig of 4714 bp in length
* 59316 59415: gap of unknown length
* 59416 65627: contig of 6212 bp in length
* 65628 65727: gap of unknown length
* 72485 72485: contig of 6758 bp in length
* 72486 72585: gap of unknown length
* 72586 77910: contig of 5325 bp in length
* 77911 78011: gap of unknown length
* 84348 84348: contig of 6338 bp in length
* 84349 84448: gap of unknown length
* 84449 90437: contig of 5989 bp in length
* 90438 90537: gap of unknown length
* 90538 97928: contig of 7391 bp in length
* 97929 98028: gap of unknown length
* 98029 105065: contig of 7037 bp in length
* 105066 105165: gap of unknown length
* 105166 111323: contig of 6158 bp in length
* 111324 111423: gap of unknown length
* 111424 117734: contig of 6311 bp in length
* 117735 117834: gap of unknown length
* 117835 125427: contig of 7593 bp in length
* 125428 125527: gap of unknown length
* 125528 134991: contig of 9464 bp in length
* 134992 135091: gap of unknown length
* 135092 143447: contig of 8356 bp in length
* 143448 143547: gap of unknown length
* 143548 156094: contig of 12547 bp in length
* 156095 156194: gap of unknown length
* 156195 157939: contig of 1745 bp in length
* 157940 158039: gap of unknown length
* 158040 175012: contig of 16972 bp in length
* 175012 175111: gap of unknown length
* 175112 187724: contig of 12613 bp in length
* 187725 187824: gap of unknown length
* 187825 208210: contig of 20386 bp in length
* 208211 208311: gap of unknown length
* 208311 213025: contig of 4715 bp in length.
* Location/Qualifiers
* 1. 213025
FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-536N17"
1. 1384
/misc_feature /note="assembly_name:Contig11"
/misc_feature /note="assembly_name:Contig14"
/misc_feature /note="assembly_name:Contig15"
/misc_feature /note="assembly_name:Contig16"
/misc_feature /note="assembly_name:Contig18"
/misc_feature /note="assembly_name:Contig19"
/misc_feature /note="assembly_name:Contig20"
/misc_feature /note="assembly_name:Contig21"
/misc_feature /note="assembly_name:Contig22"
/misc_feature /note="assembly_name:Contig23"
/misc_feature /note="assembly_name:Contig24"
/misc_feature /note="assembly_name:Contig25"
/misc_feature /note="assembly_name:Contig26"
/misc_feature /note="assembly_name:Contig27"
```

```
misc_feature 15223..17332 /note="assembly_name:Contig28"
misc_feature 17433..19823 /note="assembly_name:Contig30"
misc_feature 19924..21811 /note="assembly_name:Contig31"
misc_feature 21912..24564 /note="assembly_name:Contig32"
misc_feature 24665..27708 /note="assembly_name:Contig33"
misc_feature 27809..30176 /note="assembly_name:Contig34"
misc_feature 30277..33515 /note="assembly_name:Contig36"
misc_feature 33616..37205 /note="assembly_name:Contig37"
misc_feature 37306..39846 /note="assembly_name:Contig38"
misc_feature 39947..42384 /note="assembly_name:Contig39"
misc_feature 42485..46260 /note="assembly_name:Contig40"
misc_feature 46361..49107 /note="assembly_name:Contig41"
misc_feature /note="end:SP6"
vector:side:left"
49208..54501
/misc_feature /note="assembly_name:Contig42"
/misc_feature 54602..59315 /note="assembly_name:Contig43"
/misc_feature 59416..65627 /note="assembly_name:Contig44"
/misc_feature 65728..72485 /note="assembly_name:Contig45"
/misc_feature 72586..77910 /note="assembly_name:Contig46"
/misc_feature 78011..84348 /note="assembly_name:Contig47"
/misc_feature 84449..90437 /note="assembly_name:Contig48"
/misc_feature 90538..97928 /note="assembly_name:Contig49"
/misc_feature 98029..105065 /note="assembly_name:Contig50"
/misc_feature 105166..111323 /note="assembly_name:Contig51"
/misc_feature 111424..117734 /note="assembly_name:Contig53"
/misc_feature 117835..125427 /note="assembly_name:Contig54"
/misc_feature 125528..134991 /note="assembly_name:Contig55"
/misc_feature 135092..143447 /note="assembly_name:Contig56"
/misc_feature 143548..156094 /note="assembly_name:Contig57"
/misc_feature 156195..157939 /note="assembly_name:Contig58"
/misc_feature 158040..175011 /note="assembly_name:Contig59"
/misc_feature 175112..187724 /note="assembly_name:Contig60"
/misc_feature 187825..208210 /note="assembly_name:Contig61"
```

Query Match 12.4% Score 158; DB 2; Length 213025;
Best Local Similarity 99.5%; Pred. No. 1.5e-69;
Matches 208; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
0Y 360 TCTGGAAGTATGTCCTCAGACTTGGGAAGCCAGGCACATATTAACATA 419
DB 75909 TCTGGAAGTATGTCCTCAGACTTGGGAAGCCAGGCACATATTAACATA 75968
```

OY 420 CTGGCTGTGGTGCATGACCAATGATGTGTGAATTGGAAGCAAGTATGTG 479
|||||
DB 75969 CTGGCTGTGGTGCATGACCAATGATGTGTGAATTGGAAGCAAGTATGTG 76028
OY 480 CAAAGGTAATAATATGCGCTGGAAGTCTTAGCATATGCTTGAATGCAAGGG 539
|||||
DB 76029 CAAAGGTAATAATATGCGCTGGAAGTCTTAGCATATGCTTGAATGCAAGGG 76088
OY 540 AAACGACTGGAAGTCAATGATTAAT 568
|||||
DB 76089 AAACGACTGGAAGTCAATGATTAAT 76117
RESULT 18
AC067966 205656 bp DNA HTG 10-MAY-2000
LOCUS AC067966/c
DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-379018, *** SEQUENCING
IN PROGRESS *** 117 unordered pieces.
ACCESSION AC067966 GI:7767730
VERSION AC067966.2
KEYWORDS HTG; HTGS; PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 205656)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
JOURNAL Sequence Data
REFERENCE 2 (bases 1 to 205656)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On May 10, 2000 this sequence version replaced gi:7656685.
----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: <http://www.genomecorp.com/>
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg286
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 1008 of reads
Assembly program: Phrap; version 990315
Consensus quality: 126548 bases at least Q40
Consensus quality: 159958 bases at least Q30
Consensus quality: 170609 bases at least Q20
Insert size: 205656; sum-of-contigs
Quality coverage: 2.4x in Q20 bases; sum-of-contigs
NOTE: This is a 'working draft' sequence. It currently
* consists of 117 contigs. The true order in this sequence record is
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
I 1382: contig of 1382 bp in length
* 1383: gap of unknown length
* 2437: contig of 1055 bp in length
* 2438: gap of unknown length
* 3932: contig of 1495 bp in length
* 3933: gap of unknown length
* 5395: contig of 1463 bp in length
* 5396: gap of unknown length
* 6431: contig of 1036 bp in length
* 6432: gap of unknown length
* 7698: contig of 1267 bp in length

* 7699 8821: gap of unknown length
* 8822 9919: gap of unknown length
* 9920 11065: gap of unknown length
* 11066 12769: gap of unknown length
* 12770 14061: gap of unknown length
* 14062 15230: gap of unknown length
* 15231 16237: gap of unknown length
* 16238 17348: gap of unknown length
* 17349 18409: gap of unknown length
* 18410 19857: gap of unknown length
* 19858 21289: gap of unknown length
* 21290 22433: gap of unknown length
* 22434 23819: gap of unknown length
* 23820 25788: gap of unknown length
* 25789 26867: gap of unknown length
* 26868 28331: gap of unknown length
* 28332 29340: gap of unknown length
* 29341 30459: gap of unknown length
* 30460 31600: gap of unknown length
* 31601 32790: gap of unknown length
* 32791 33825: gap of unknown length
* 33826 35833: gap of unknown length
* 35834 37115: gap of unknown length
* 37116 38775: gap of unknown length
* 38776 40065: gap of unknown length
* 40066 41910: gap of unknown length
* 41911 44094: gap of unknown length
* 44095 45463: gap of unknown length
* 45464 46887: gap of unknown length
* 46888 48429: gap of unknown length
* 48430 49705: gap of unknown length
* 49706 51649: gap of unknown length
* 51650 52985: gap of unknown length
* 52986 54144: gap of unknown length
* 54145 55442: gap of unknown length
* 55443 57106: gap of unknown length

Query Match 10.28; Score 130; DB 2; Length 205656;
Best Local Similarity 100.0%; Pred. No. 3; 2e-55;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

57107 58300: contig of 1194 bp in length
58301 59380: contig of 1080 bp in length
59381 60547: contig of 1167 bp in length
60548 62394: contig of 1847 bp in length
62395 63433: contig of 1039 bp in length
63434 64990: contig of 1557 bp in length
64991 66406: contig of 1416 bp in length
66407 67850: contig of 1444 bp in length
67851 69251: contig of 1401 bp in length
69252 70304: contig of 1053 bp in length
70305 72354: contig of 2050 bp in length
72355 74094: contig of 1740 bp in length
74095 75933: contig of 1839 bp in length
75934 77482: contig of 1549 bp in length
77483 78801: contig of 1319 bp in length
78802 80209: contig of 1408 bp in length
80210 81738: contig of 1529 bp in length
81739 83669: contig of 1931 bp in length
83670 84875: contig of 1206 bp in length
84876 86866: contig of 1811 bp in length
86867 87717: contig of 1031 bp in length
87718 88952: contig of 1235 bp in length
88953 90907: contig of 1955 bp in length
90908 92344: contig of 1437 bp in length
92345 93840: contig of 1496 bp in length
93841 95003: contig of 1163 bp in length
95004 97207: contig of 2204 bp in length
97208 99107: contig of 1900 bp in length
99108 100938: contig of 1831 bp in length
100939 103110: contig of 2172 bp in length
103111 105202: contig of 2092 bp in length
105203 106758: contig of 1556 bp in length
106759 108102: contig of 1344 bp in length
108103 109358: contig of 1256 bp in length
109359 110883: contig of 1525 bp in length

OY 599 GATATCAATGATGTCAAAGCGCTGAACCTGCTACTAGAGCTACTGCTGTT 658
|||||
Db 20249 GATATCAATGATGTCAAAGCGCTGAACCTGCTACTAGAGCTACTGCTGTT 20190
OY 659 AGAAGTATAGTTCCTGATGAGAAACAGAAATGATGCTGTTAATGAGAAATT 718
|||||
Db 20189 AGAAGTATAGTTCCTGATGAGAAACAGAAATGATGCTGTTAATGAGAAATT 20130
OY 719 AAAGATAGG 728
|||||
Db 20129 AAAGATAGG 20120

RESULT 19
AC067746
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

AC067746 195126 bp DNA HTG 16-MAY-2001
Homo sapiens chromosome 10 clone RP11-408L20, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC067746
AC067746.5 GI:14091814
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 195126)
Smith,D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
2 (bases 1 to 195126)
Smith,D.R.
Direct Submission
Submitted (27-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On May 16, 2001 this sequence version replaced gi:8389438.
Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
Project Information
Center project name: hg288
Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 188219 bases at least Q40
Consensus quality: 189837 bases at least Q30
Consensus quality: 190569 bases at least Q20
Insert size: 194225; sum-of-contigs
Quality coverage: 6.7x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1102: contig of 1102 bp in length
1103 1202: gap of unknown length
1203 2394: contig of 1192 bp in length
2395 2494: gap of unknown length
2495 3744: contig of 1250 bp in length
3745 3844: gap of unknown length
3845 5030: contig of 1186 bp in length
5031 5130: gap of unknown length
5131 15992: contig of 10862 bp in length
15993 16093: gap of unknown length
16093 31383: contig of 15291 bp in length

FEATURES
source
1. 195126
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-408L20"
/clone_lib="RP11-11"
1. 1102
/note="assembly_name:Contig16"
1203. 2394
/note="assembly_name:Contig18"
2495. 3744
/note="assembly_name:Contig19"
3845. 5030
/note="assembly_name:Contig21"
5131. 15992
/note="assembly_name:Contig31"
16093. 31383
/note="assembly_name:Contig32"
31484. 47125
/note="assembly_name:Contig33"
clone_end="17"
47226. 69920
/note="assembly_name:Contig34"
70021. 92732
/note="assembly_name:Contig35"
clone_end="SP6"
92833. 118005
/note="assembly_name:Contig36"
118106. 195126
/note="assembly_name:Contig37"
BASE COUNT 52995 a 46055 c 45277 g 49797 t 1002 others
ORIGIN
Query Match 7.2%: Score 92; DB 2; Length 195126;
Best Local Similarity 100.0%; Pred. No. 8.5e-36;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 384 AGACTGGGAAGACCCAGGCGACATGATTAACATGCTGCTGTGTGATGACATGACC 443
DB 2533 AGACTGGGAAGACCCAGGCGACATGATTAACATGCTGCTGTGTGATGACATGACC 2592
QY 444 CAATTGATGTGTGTAATTGAATGGAACAGTA 475
DB 2593 CAATTGATGTGTGTAATTGAATGGAACAGTA 2624
RESULT 20
AC067749/c DNA HTG 26-MAY-2001
LOCUS AC067749 215049 bp
DEFINITION Homo sapiens chromosome 10 clone RP11-367H5, WORKING DRAFT
SEQUENCE 33 unordered pieces.
AC067749
VERSION AC067749.4 GI:14209706
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 215049)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Sequence Data
Unpublished
2 (bases 1 to 215049)
Smith,D.R.
Direct Submission
Submitted (27-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On May 26, 2001 this sequence version replaced gi:8954013.
Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
Project Information
Center project name: hg285
Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 194485 bases at least Q40
Consensus quality: 199162 bases at least Q30
Consensus quality: 201282 bases at least Q20
Insert size: 211948; sum-of-Contigs
Quality coverage: 6.8x in Q20 bases; sum-of-Contigs
NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1163: contig of 1163 bp in length
1164: gap of unknown length
1263: contig of 1013 bp in length
1264: gap of unknown length
2276: gap of unknown length
2377: contig of 1079 bp in length
3455: gap of unknown length
3555: gap of 1032 bp in length
4587: gap of unknown length
4688: gap of 1008 bp in length
5695: gap of unknown length
5795: gap of 1175 bp in length
6970: contig of 1175 bp in length
6971: gap of unknown length
7071: gap of 1087 bp in length
8157: contig of 1087 bp in length
8257: gap of unknown length
8258: contig of 1043 bp in length
9300: gap of unknown length
9301: gap of unknown length
10400: gap of 1041 bp in length
10441: contig of 1041 bp in length
10442: gap of unknown length
10541: gap of unknown length
1163: contig of 1122 bp in length
1164: gap of unknown length
1165: gap of unknown length
12926: contig of 1163 bp in length
13026: gap of unknown length
13027: gap of 1142 bp in length
14168: contig of 1142 bp in length
14268: gap of unknown length
14269: gap of 1306 bp in length
15574: contig of 1306 bp in length
15674: gap of unknown length
1575: gap of 1364 bp in length
1675: gap of unknown length
17038: contig of 1364 bp in length
17138: gap of unknown length
17139: contig of 1263 bp in length
18401: gap of unknown length
18501: gap of 1287 bp in length
19788: contig of 1287 bp in length
19888: gap of unknown length
21245: contig of 157 bp in length
21246: gap of unknown length
21345: gap of 1722 bp in length
23067: contig of 1722 bp in length
23167: gap of unknown length
25167: contig of 2000 bp in length
25267: gap of unknown length
28266: contig of 2399 bp in length
28366: gap of unknown length

```
* 28367 30737: contig of 2371 bp in length
* 30738 30837: gap of unknown length
* 30838 35830: contig of 4993 bp in length
* 35831 35930: gap of unknown length
* 43874 43973: contig of 7943 bp in length
* 43974 43974: gap of unknown length
* 54314 54313: contig of 10340 bp in length
* 54414 62524: gap of unknown length
* 62525 62524: gap of unknown length
* 75785 75784: contig of 13160 bp in length
* 75785 92445: gap of unknown length
* 92446 92445: contig of 16561 bp in length
* 92546 109803: contig of 17258 bp in length
* 109804 109903: gap of unknown length
* 109904 128494: contig of 18591 bp in length
* 128495 128594: gap of unknown length
* 128595 146739: contig of 18145 bp in length
* 146740 146839: gap of unknown length
* 167230 167330: contig of 20391 bp in length
* 167231 167331: gap of unknown length
* 187365 187365: contig of 20035 bp in length
* 187366 187465: gap of unknown length
* 187466 215049: contig of 27584 bp in length.
* Location/Qualifiers
  1. 215049
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="10"
  /clone_id="RP11-367H5"
  /clone_id="RP11-367H5"
  1. 1163
  /note="assembly_name:Contig4"
  1264. 2276
  /note="assembly_name:Contig4"
  2377. 3435
  /note="assembly_name:Contig12"
  3556. 4587
  /note="assembly_name:Contig15"
  4688. 5695
  /note="assembly_name:Contig55"
  5796. 6970
  /note="assembly_name:Contig60"
  7071. 8157
  /note="assembly_name:Contig62"
  8258. 9300
  /note="assembly_name:Contig63"
  9401. 10441
  /note="assembly_name:Contig69"
  10542. 11663
  /note="assembly_name:Contig73"
  11764. 12926
  /note="assembly_name:Contig75"
  13027. 14168
  /note="assembly_name:Contig112"
  14269. 15574
  /note="assembly_name:Contig113"
  15675. 17038
  /note="assembly_name:Contig116"
  17139. 18401
  /note="assembly_name:Contig118"
  18502. 19788
  /note="assembly_name:Contig121"
  19889. 21245
  /note="assembly_name:Contig124"
  21346. 23067
  /note="assembly_name:Contig127"
  23168. 25167
  /note="assembly_name:Contig134"
  clone_end:596"
  misc_feature
  25268. 28266
  /note="assembly_name:Contig135"
```

```
misc_feature 28367. 30737
/note="assembly_name:Contig136"
misc_feature 30838. 35830
/note="assembly_name:Contig137"
misc_feature 35931. 43873
/note="assembly_name:Contig138"
misc_feature 43974. 54313
/note="assembly_name:Contig139"
misc_feature 54414. 62524
/note="assembly_name:Contig140"
misc_feature 62625. 75784
/note="assembly_name:Contig141"
misc_feature 75885. 92445
/note="assembly_name:Contig142"
misc_feature 92546. 109803
/note="assembly_name:Contig143"
misc_feature 109904. 128494
/note="assembly_name:Contig144"
misc_feature 128595. 146739
/note="assembly_name:Contig145"
misc_feature 146840. 167230
/note="assembly_name:Contig146"
misc_feature 167331. 187365
/note="assembly_name:Contig147"
misc_feature 187466. 215049
/note="assembly_name:Contig148"
clone_end:"7"

BASE COUNT 57394 a 47838 c 47126 g 59466 t 3225 others
ORIGIN
```

```
Query Match 7.2%; Score 92; DB 2; Length 215049;
Best Local Similarity 100.0%; Pred. No. 8.56-36;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 599 GATATCATGATGCTCAACGCTGAACTGCTACTTGAAGCTACTGTGACACTGTTT 658
|||||
Db 47543 GATATCAATGATGCTCAACGCTGAACTGCTACTTGAAGCTACTGTGACACTGTTT 47484
|||||
QY 659 ACAAGGTATTAAGTTCCTGATGTAAGCAAAATGATTTGGCTTAATGACAAATT 718
|||||
Db 47483 ACAAGGTATTAAGTTCCTGATGTAAGCAAAATGATTTGGCTTAATGACAAATT 47424
|||||
QY 719 AAAGATTAAGG 728
|||||
Db 47423 AAAGATTAAGG 47414
```

```
RESULT 21
AC011012 155645 bp DNA HTG 03-NOV-2000
LOCUS Homo sapiens clone RP11-8023, WORKING DRAFT SEQUENCE, 7 unordered
DEFINITION pieces.
AC011012
AC011012.4 GI:11079562
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155645)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Baines,N., Beckwith,J., Birney,D., Bork,P., Bouckhwalter,B.,
Brown,A., Castle,A., Collins,S., Collins,S., Collins,S., Collins,S.,
Cooke,P., Dearfello,K., Dewar,K., Domino,M., Donlan,L., Doyle,M.,
Ferrelia,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gaidyna,S., Grant,G., Hargis,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Margulis,N.,
```

TITLE
JOURNAL
COMMENT

McEwan, P., McEwan, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomas, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testa, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 3, 2000 this sequence version replaced g1:7239549.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
Project information

Center project name: L2966
Center clone name: 8.D.23

Summary Statistics
Sequencing vector: M13; M7815; 55% of reads
Sequencing vector: Plasmid; n/a; 45% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 153969 bases at least Q40
Consensus quality: 154547 bases at least Q30
Consensus quality: 154815 bases at least Q20
Insert size: 152000; agarose-fp
Insert size: 155045; sum-of-contents
Quality coverage: 12.1 in Q20 bases; agarose-fp
Quality coverage: 11.9 in Q20 b.

NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 142: contig of 142 bp in length
143 242: gap of 100 bp
243 889: contig of 647 bp in length
890 989: gap of 100 bp
990 1642: contig of 653 bp in length
1643 1742: gap of 100 bp
1743 2403: contig of 661 bp in length
2404 2503: gap of 100 bp
2504 3180: contig of 677 bp in length
3181 3280: gap of 100 bp
3281 4042: contig of 762 bp in length
4043 4142: gap of 100 bp
4143 155645: contig of 151503 bp in length.

FEATURES

SOURCE

1. 155645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-8D23"
/clone_1lb="RP11-11 Human Male BAC"
1. 142
/note="assembly_fragment"
243. 889
/note="assembly_fragment"
990. 1642
/note="assembly_fragment"
1743. 2403
/note="assembly_fragment"
2504. 3180
/note="assembly_fragment"
3281. 4042
/note="assembly_fragment"
4143. 155645
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

BASE COUNT 49201 a 32211 c 29880 g 43751 t 602 others
ORIGIN

Query Match 6.2%; Score 79; DB 2; Length 155645;
Best Local Similarity 100.0%; Pred. No. 3.8e-29;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 834 AGAGCCCTTCAAGTGTATCTGATGCTGCGACAGACCATTTGGATGCTTTACACAC 893
|||||
DB 7656 AGAGCCCTTCAAGTGTATCTGATGCTGCGACAGACCATTTGGATGCTTTACACAC 76645
|||||

QY 894 CCTGTGAATCTGCTGCAC 912.
|||||
DB 76646 CCTGTGAATCTGCTGCAC 76664

RESULT 22

AC025103 70575 bp DNA HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-25A15, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC025103.1 GI:7158913
ACCESSION AC025103.1 GI:7158913
VERSION HTG; HTGS_PHASE0.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 70575)
Britten, B., Linton, L., Nussbaum, C. and Lander, E.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 70575)
Britten, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, T., Bastien, V., Bede, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Chopel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McEwan, A., McKernan, K., McPheters, R.,
Meldrum, J., Menes, L., Mihova, T., Miranda, C., Miñana, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, L.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomas, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
Project information
Center project name: L4527
Center clone name: 25_A.15

* NOTE: This record contains 75 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 849: contig of 849 bp in length
850 949: gap of 100 bp
950 1823: contig of 874 bp in length
1824 1923: gap of 100 bp
1924 2761: contig of 838 bp in length
2762 2861: gap of 100 bp
2862 3738: contig of 877 bp in length
3739 3838: gap of 100 bp
3839 4684: contig of 846 bp in length
4685 4784: gap of 100 bp
4785 5604: contig of 820 bp in length
5605 5704: gap of 100 bp
5705 6518: contig of 814 bp in length
6519 6618: gap of 100 bp
6619 7445: contig of 827 bp in length
7446 7545: gap of 100 bp
7546 8391: contig of 846 bp in length
8392 8491: gap of 100 bp
8492 9312: contig of 821 bp in length
9313 9412: gap of 100 bp
9413 10228: contig of 816 bp in length
10229 10328: gap of 100 bp
10329 11170: contig of 842 bp in length
11171 11270: gap of 100 bp
11271 12097: contig of 827 bp in length
12098 12197: gap of 100 bp
12198 13031: contig of 834 bp in length
13032 13131: gap of 100 bp
13132 13940: contig of 809 bp in length
13941 14040: gap of 100 bp
14041 14887: contig of 847 bp in length
14888 14987: gap of 100 bp
14988 15823: contig of 836 bp in length
15824 15923: gap of 100 bp
15924 16748: contig of 825 bp in length
16749 16848: gap of 100 bp
16849 17692: contig of 844 bp in length
17693 17792: gap of 100 bp
17793 18636: contig of 844 bp in length
18637 18736: gap of 100 bp
18737 19596: contig of 860 bp in length
19597 19696: gap of 100 bp
19697 20534: contig of 838 bp in length
20535 20634: gap of 100 bp
20635 21446: contig of 812 bp in length
21447 21546: gap of 100 bp
21547 22391: contig of 845 bp in length
22392 22491: gap of 100 bp
22492 23319: contig of 828 bp in length
23320 23419: gap of 100 bp
23420 24268: contig of 849 bp in length
24269 24368: gap of 100 bp
24369 25218: contig of 850 bp in length
25219 25318: gap of 100 bp
25319 26161: contig of 843 bp in length
26162 26261: gap of 100 bp
26262 27120: contig of 859 bp in length
27121 27220: gap of 100 bp
27221 28039: contig of 819 bp in length
28040 28139: gap of 100 bp
28140 28978: contig of 839 bp in length
28979 29078: gap of 100 bp
29079 29937: contig of 859 bp in length
29938 30037: gap of 100 bp
30038 30860: contig of 823 bp in length

30861 30960: gap of 100 bp
30961 31777: contig of 817 bp in length
31778 31877: gap of 100 bp
31878 32749: contig of 872 bp in length
32750 32849: gap of 100 bp
32850 33680: contig of 831 bp in length
33681 33780: gap of 100 bp
33781 34625: contig of 845 bp in length
34626 34725: gap of 100 bp
34726 35561: contig of 836 bp in length
35562 35661: gap of 100 bp
35662 36540: contig of 879 bp in length
36541 36640: gap of 100 bp
36641 37473: contig of 833 bp in length
37474 37573: gap of 100 bp
37574 38378: contig of 805 bp in length
38379 38478: gap of 100 bp
38479 39345: contig of 867 bp in length
39346 39445: gap of 100 bp
39446 40238: contig of 793 bp in length
40239 40338: gap of 100 bp
40339 41169: contig of 831 bp in length
41170 41269: gap of 100 bp
41270 42089: contig of 820 bp in length
42090 42189: gap of 100 bp
42190 43054: contig of 865 bp in length
43055 43154: gap of 100 bp
43155 44013: contig of 859 bp in length
44014 44113: gap of 100 bp
44114 44982: contig of 869 bp in length
44983 45082: gap of 100 bp
45083 45956: contig of 874 bp in length
45957 46056: gap of 100 bp
46057 46687: contig of 831 bp in length
46688 46987: gap of 100 bp
46988 47800: contig of 813 bp in length
47801 47900: gap of 100 bp
47901 48759: contig of 859 bp in length
48760 48859: gap of 100 bp
48860 49719: contig of 860 bp in length
49720 49819: gap of 100 bp
49820 50690: contig of 871 bp in length
50691 50790: gap of 100 bp
50791 51641: contig of 851 bp in length
51642 51741: gap of 100 bp
51742 52602: contig of 861 bp in length
52603 52702: gap of 100 bp
52703 53577: contig of 875 bp in length
53578 53677: gap of 100 bp
53678 54522: contig of 845 bp in length
54523 54622: gap of 100 bp
54623 55470: contig of 848 bp in length
55471 55570: gap of 100 bp
55571 56408: contig of 838 bp in length
56409 56508: gap of 100 bp
56509 57360: contig of 852 bp in length
57361 57460: gap of 100 bp
57461 58286: contig of 826 bp in length
58287 58386: gap of 100 bp
58387 59224: contig of 838 bp in length
59225 59324: gap of 100 bp
59325 60131: contig of 807 bp in length
60132 60231: gap of 100 bp
60232 61082: contig of 851 bp in length
61083 61182: gap of 100 bp
61183 62027: contig of 845 bp in length
62028 62127: gap of 100 bp
62128 62985: contig of 858 bp in length
62986 63085: gap of 100 bp
63086 63913: contig of 828 bp in length
63914 64013: gap of 100 bp
64014 64873: contig of 860 bp in length
64874 64973: gap of 100 bp

* 64974 65814: contig of 841 bp in length

Query Match 5.8%; Score 74; DB 2; Length 70575;

Best Local Similarity 100.0%; Pred. No. 1.4e-26;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 425 TGTGTGTCGACATGACCAATGATGTCGTGTAATGGAAGCAAGGTATGTCAGCA 484

DB 33900 TGTGTGTCGACATGACCAATGATGTCGTGTAATGGAAGCAAGGTATGTCAGCA 33959

OY 485 GGTGAATTAATTGG 498

DB 33960 GGTGAATTAATTGG 33973

RESULT 23 BOVIOPPP 1266 bp mRNA MAM 27-APR-1993

LOCUS BOVIOPPP Bovine inorganic pyrophosphatase mRNA sequence.

DEFINITION M95283

VERSION M95283.1 GI:163228

KEYWORDS Inorganic pyrophosphatase.

SOURCE Bos taurus retina cDNA to mRNA.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 1266)

AUTHORS Yang, Z. and Wensel, T. G.

TITLE Molecular cloning and functional expression of cDNA encoding a

JOURNAL mammalian inorganic pyrophosphatase

FEATURES J. Biol. Chem. 267, 24641-24647 (1992)

Source Location/Qualifiers

1..1266

/organism="Bos taurus"

/db_xref="taxon:9913"

/tissue_type="retina"

BASE COUNT 396 a 239 c 300 g 331 t

ORIGIN

Query Match 4.4%; Score 56; DB 4; Length 1266;

Best Local Similarity 100.0%; Pred. No. 2.6e-17;

Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 419 ACTGGCTGTGTGTCGACCAATGACCAATGATGTCGTGTAATGGAAGCAAGGT 474

DB 399 ACTGGCTGTGTGTCGACCAATGACCAATGATGTCGTGTAATGGAAGCAAGGT 454

RESULT 24 BC010468 1243 bp mRNA ROD 12-JUL-2001

LOCUS BC010468 Mus musculus, RIKEN cDNA 2010317E03 gene, clone MGC:6716

DEFINITION IMAGE:3585780, mRNA, complete cds.

ACCESSION BC010468

VERSION BC010468.1 GI:14714656

KEYWORDS MGC.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

REFERENCE 1 (bases 1 to 1243)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgaps-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalona@bcm.tmc.edu

Villalona, D.K., Luna, R.A., Hale, S.M., Huylk, S., Lu, X., Garcia,

A.M., Hollway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,

Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/HLN at: <http://image.llnl.gov>

Series: IRAC Plate: 10 Row: f Column: 15

This clone was selected for full length sequencing because it

passed the following selection criteria: Genomescan gene

prediction, Similarity but not identity to protein.

Location/Qualifiers

1..1243

/organism="Mus musculus"

/db_xref="LocusID:67895"

/db_xref="taxon:10090"

/clone="MGC:6716 IMAGE:3585780"

/tissue_type="Mammary tumor, Metallothionin-TGF alpha

model. 10 month old virgin mouse. Taken by biopsy."

/clone_id="NCI_CGAP_Mam1"

/lab_host="DH10B"

/note="Vector: PCMV-SPORT6"

42..911

/codon_start=1

/product="Riken cDNA 2010317E03 gene"

/protein_id="AAH10468.1"

/db_xref="GI:14714657"

/translation="MSGFSSSEERAAPTLEYRFLKNEGQVYSPHVDPIADKDF

HMVEVPRMSNAKKEIATKDPINIKODVAKIYVYANLPYKGIWNGAIPQWE

DPGSDKHTGCCGNDPILDCVETSKYCARGEIIRVGLILAMIDGCTMDVIAIN

VDDPDANIKDISVPELAKREYILEATDMDPRKRVKVPGRKRENFAPFAEKKNDFAVD

IKSTHYRWALVTKTDGKIGISCMNTTVSESPKCPDPAKAIYDALPPCESACSL

PFDVDMFHDOKN"

BASE COUNT 366 a 270 c 310 g 297 t

ORIGIN

Query Match 3.9%; Score 50; DB 10; Length 1243;

Best Local Similarity 100.0%; Pred. No. 3e-14;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 653 TGGTTAGAGGTATAGGTTCTGATGGAAGCAAGAAATGAGTTGC 702

DB 606 TGGTTAGAGGTATAGGTTCTGATGGAAGCAAGAAATGAGTTGC 655

RESULT 25 AC025103 70575 bp DNA HTG 13-JUL-2000

LOCUS AC025103 Homo sapiens clone RP11-25A15, LOW-PASS SEQUENCE SAMPLING.

DEFINITION AC025103

ACCESSION AC025103.1 GI:7158913

VERSION HTG; HTGS_PHASE0.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 70575)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Unpublished

JOURNAL 2 (bases 1 to 70575)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Balwin, J., Barina, N., Bastien, V., Bida, F.,

Boguslavsky, L., Bouknight, B., Brown, A., Burkett, G.,

Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,

TITLE
JOURNAL
COMMENT

Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardina, S., Glend, S., Goeyre, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehocsky, J.,
Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienda, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisanal, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, W.

Direct Submission
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L4527
Center clone name: 25_A15

NOTE: This record contains 75 individual
sequencing reads that have not been assembled into
contigs. Runs of N are used to separate the reads
and the order in which they appear is completely
arbitrary. Low-pass sequence sampling is useful for
identifying clones that may be gene-rich and allows
overlap relationships among clones to be deduced.
However, it should not be assumed that this clone
will be sequenced to completion. In the event that
the record is updated, the accession number will
be preserved.

1 849: contig of 849 bp in length
850 949: gap of 100 bp
950 1823: contig of 874 bp in length
1824 1923: gap of 100 bp
1924 2761: contig of 838 bp in length
2762 2861: gap of 100 bp
2862 3738: contig of 877 bp in length
3739 3838: gap of 100 bp
3839 4684: contig of 846 bp in length
4685 4784: gap of 100 bp
4785 5604: contig of 820 bp in length
5605 5704: gap of 100 bp
5705 6518: contig of 814 bp in length
6519 6618: gap of 100 bp
6619 7445: contig of 827 bp in length
7446 7545: gap of 100 bp
7546 8391: contig of 846 bp in length
8392 8491: gap of 100 bp
8492 9312: contig of 821 bp in length
9313 9412: gap of 100 bp
9413 10228: contig of 816 bp in length
10229 10328: gap of 100 bp
10329 11170: contig of 842 bp in length
11171 11270: gap of 100 bp
11271 12097: contig of 827 bp in length
12098 12197: gap of 100 bp
12198 13031: contig of 834 bp in length
13032 13131: gap of 100 bp
13132 13940: contig of 809 bp in length
13941 14040: gap of 100 bp

14041 14887: contig of 847 bp in length
14888 14987: gap of 100 bp
14988 15823: contig of 836 bp in length
15824 15923: gap of 100 bp
15924 16748: contig of 825 bp in length
16749 16848: gap of 100 bp
16849 17692: contig of 844 bp in length
17693 17792: gap of 100 bp
17793 18636: contig of 844 bp in length
18637 18736: gap of 100 bp
18737 18596: contig of 860 bp in length
18597 19696: gap of 100 bp
19697 20534: contig of 838 bp in length
20535 20634: gap of 100 bp
20635 21446: contig of 812 bp in length
21447 21546: gap of 100 bp
21547 22391: contig of 845 bp in length
22392 22491: gap of 100 bp
22492 23319: contig of 828 bp in length
23320 23419: gap of 100 bp
23420 24268: contig of 849 bp in length
24269 24368: gap of 100 bp
24369 25218: contig of 830 bp in length
25219 25318: gap of 100 bp
25319 26161: contig of 843 bp in length
26162 26261: gap of 100 bp
26262 27120: contig of 859 bp in length
27121 27220: gap of 100 bp
27221 28039: contig of 819 bp in length
28040 28139: gap of 100 bp
28140 28978: contig of 839 bp in length
28979 29078: gap of 100 bp
29079 29937: contig of 859 bp in length
29938 30037: gap of 100 bp
30038 30860: contig of 823 bp in length
30861 30960: gap of 100 bp
30961 31777: contig of 817 bp in length
31778 31877: gap of 100 bp
31878 32748: contig of 872 bp in length
32750 32848: gap of 100 bp
32850 33680: contig of 831 bp in length
33681 33780: gap of 100 bp
33781 34625: contig of 845 bp in length
34626 34725: gap of 100 bp
34726 35561: contig of 836 bp in length
35562 35661: gap of 100 bp
35662 36540: contig of 879 bp in length
36541 36640: gap of 100 bp
36641 37473: contig of 833 bp in length
37474 37573: gap of 100 bp
37574 38378: contig of 865 bp in length
38379 38478: gap of 100 bp
38479 39345: contig of 867 bp in length
39346 39445: gap of 100 bp
39446 40238: contig of 793 bp in length
40239 40338: gap of 100 bp
40339 41169: contig of 831 bp in length
41170 41269: gap of 100 bp
41270 42089: contig of 820 bp in length
42090 42189: gap of 100 bp
42190 43054: contig of 865 bp in length
43055 43154: gap of 100 bp
43155 44013: contig of 859 bp in length
44014 44113: gap of 100 bp
44114 44982: contig of 869 bp in length
44983 45082: gap of 100 bp
45083 45956: contig of 874 bp in length
45957 46056: gap of 100 bp
46057 46887: contig of 831 bp in length
46888 46987: gap of 100 bp
46988 47800: contig of 813 bp in length
47801 47900: gap of 100 bp
47901 48759: contig of 859 bp in length

```

* 48760 48859: gap of 100 bp
* 48860 49719: contig of 860 bp in length
* 49720 49819: gap of 100 bp
* 49820 50690: contig of 871 bp in length
* 50691 50790: gap of 100 bp
* 50791 51641: contig of 851 bp in length
* 51642 51741: gap of 100 bp
* 51742 52602: contig of 861 bp in length
* 52603 52702: gap of 100 bp
* 52703 53577: contig of 875 bp in length
* 53578 53677: gap of 100 bp
* 53678 54522: contig of 845 bp in length
* 54523 54622: gap of 100 bp
* 54623 55470: contig of 848 bp in length
* 55471 55570: gap of 100 bp
* 55571 56408: contig of 838 bp in length
* 56409 56508: gap of 100 bp
* 56509 57360: contig of 852 bp in length
* 57361 57460: gap of 100 bp
* 57461 58286: contig of 826 bp in length
* 58287 58386: gap of 100 bp
* 58387 59224: contig of 838 bp in length
* 59225 59324: gap of 100 bp
* 59325 60131: contig of 807 bp in length
* 60132 60231: gap of 100 bp
* 60232 61082: contig of 851 bp in length
* 61083 61182: gap of 100 bp
* 61183 62027: contig of 845 bp in length
* 62028 62127: gap of 100 bp
* 62128 62985: contig of 858 bp in length
* 62986 63085: gap of 100 bp
* 63086 63913: contig of 828 bp in length
* 63914 64013: gap of 100 bp
* 64014 64873: contig of 860 bp in length
* 64874 64973: gap of 100 bp
* 64974 65814: contig of 841 bp in length

```

Query Match 3.5%; Score 44; DB 2; Length 70575;
 Best Local Similarity 100.0%; Pred. No. 3.1e-11;
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 GAAACTTCGCTAGTTCGCAATTTGTCGCCGATTAAGGATAT 358
 Db 48087 GAAACTTCGCTAGTTCGCAATTTGTCGCCGATTAAGGATAT 48044

RESULT 26
 AL513487/c AL513487 144177 bp DNA PRI 09-MAR-2001
 LOCUS Human DNA sequence from clone RP13-63115 on chromosome Xq24-26.2,
 DEFINITION complete sequence.
 ACCESSION AL513487
 VERSION AL513487.3 GI:13274904
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 144177)
 AUTHORS Heath, P.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Mar 12, 2001 this sequence version replaced gi:12831924.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate

FEATURES

source

chemistry or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/chrX> RP13-63115 is from the library RP13-13.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6

This sequence is the entire insert of clone RP13-63115.

Location/Qualifiers

1..144177

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/map="q24-26.2"

/clone="RP13-63115"

/clone_id="RP13-13.1"

12..660

/note="LIM2 repeat: matches 1189..1850 of consensus"

703..1121

/note="Lip repeat: matches 1681..2103 of consensus"

1193..1343

/note="LIP15-16 repeat: matches 1023..1176 of consensus"

1350..1822

/note="LIP15-16 repeat: matches 425..886 of consensus"

2031..2304

/note="LIP15-16 repeat: matches 109..378 of consensus"

2664..2875

/note="LIP12 repeat: matches -1422..-1209 of consensus"

2896..3326

/note="LIMc repeat: matches 1398..1849 of consensus"

3328..3351

/note="12 copies 2 mer tg 95% conserved"

3353..3380

/note="14 copies 2 mer ac 100% conserved"

3422..3693

/note="AluSc repeat: matches 1..273 of consensus"

3694..4219

/note="LIMc repeat: matches 844..1412 of consensus"

4233..4375

/note="THEIB repeat: matches 3..159 of consensus"

4374..4470

/note="THEIB repeat: matches 49..159 of consensus"

4471..5423

/note="THEIB-INTERNAL repeat: matches 610..1571 of consensus"

5424..5786

/note="THEIB repeat: matches 2..364 of consensus"

5863..5983

/note="LIMc repeat: matches 793..913 of consensus"

6009..6319

/note="AluSc repeat: matches 1..312 of consensus"

6824..6882

/note="LIMc repeat: matches 11..67 of consensus"

7277..10034

/note="LIP15 repeat: matches 3280..6157 of consensus"

10035..10343

/note="Alu repeat: matches 1..302 of consensus"

10344..12265

/note="LIP15 repeat: matches 1412..3280 of consensus"

14439..14804

/note="THEIB repeat: matches 1..364 of consensus"

14805..16355


```

/note="THE1B-INTERNAL repeat: matches 1. .1580 of
consensus"
repeat_region
16356. .16721
/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region
17289. .17320
/note="16 copies 2 mer tt 84% conserved"
repeat_region
18611. .18725
/note="L1M4 repeat: matches 4363. .4484 of consensus"
repeat_region
20495. .20649
/note="L1P16 repeat: matches 4305. .4448 of consensus"
repeat_region
20702. .22403
/note="L1P16 repeat: matches 4444. .6157 of consensus"
repeat_region
22625. .22878
/note="MIR repeat: matches 13. .262 of consensus"
repeat_region
24431. .24916
/note="L2 repeat: matches 1778. .2276 of consensus"
repeat_region
25602. .26550
/note="L2 repeat: matches 1744. .2741 of consensus"
repeat_region
26745. .27153
/note="L1P repeat: matches 1450. .1858 of consensus"
repeat_region
27138. .28917
/note="L1P3 repeat: matches 4051. .5838 of consensus"
repeat_region
28918. .29237
/note="L1P3 repeat: matches 5826. .6146 of consensus"
repeat_region
29259. .33467
/note="L1P8 repeat: matches 1843. .6163 of consensus"
repeat_region
33618. .33724
/note="L1M4A repeat: matches 6194. .6298 of consensus"
repeat_region
34017. .34962
/note="L1P82 repeat: matches 5195. .6155 of consensus"
repeat_region
35433. .35654
/note="MIR repeat: matches 22. .249 of consensus"
repeat_region
35691. .35757
/note="L1M8 repeat: matches 6099. .6170 of consensus"
repeat_region
36190. .36444
/note="L1P2 repeat: matches 5892. .6146 of consensus"
repeat_region
37225. .37318
/note="47 copies 2 mer ta 68% conserved"
repeat_region
38864. .39056
/note="MIR repeat: matches 42. .252 of consensus"
repeat_region
39283. .39349
/note="L2 repeat: matches 2595. .2668 of consensus"
repeat_region
39591. .41301
/note="L1M2 repeat: matches 4395. .6137 of consensus"
repeat_region
41295. .42595
/note="L1P3 repeat: matches 4770. .6117 of consensus"
repeat_region
42611. .42634
/note="L1M2 repeat: matches 4374. .4396 of consensus"
repeat_region
42635. .43740
/note="MER1C repeat: matches 3. .1071 of consensus"
repeat_region
43741. .44993
/note="L1M2 repeat: matches 3112. .4374 of consensus"
repeat_region
45006. .45402
/note="L1P7 repeat: matches 5152. .5564 of consensus"
repeat_region
45403. .45973
/note="L1P7 repeat: matches 5552. .6141 of consensus"
repeat_region
46308. .46671
/note="L1P16 repeat: matches 5824. .6150 of consensus"
repeat_region
47439. .47743
/note="AluY repeat: matches 1. .304 of consensus"
repeat_region
49898. .50115
/note="MIR repeat: matches 33. .260 of consensus"
repeat_region
50777. .51235
/note="L1M9 repeat: matches 5837. .6306 of consensus"
repeat_region
52537. .52671
/note="L1M7 repeat: matches 6149. .6284 of consensus"
repeat_region
52998. .53083
/note="L2 repeat: matches 2295. .2380 of consensus"
repeat_region
53176. .53286
/note="L2 repeat: matches 2604. .2716 of consensus"
repeat_region
54010. .54049
/note="20 copies 2 mer ca 100% conserved"
repeat_region
54055. .54082

```

```

/note="14 copies 2 mer ag 100% conserved"
repeat_region
54126. .54204
/note="M4E1 repeat: matches 2. .80 of consensus"
repeat_region
54804. .54897
/note="M4E3 repeat: matches 1. .90 of consensus"
repeat_region
55936. .56044
/note="MIR repeat: matches 35. .150 of consensus"
repeat_region
56467. .56656
/note="M4E3 repeat: matches 1. .209 of consensus"
repeat_region
57324. .57446
/note="L1M4C repeat: matches 1927. .2055 of consensus"
repeat_region
58142. .58455
/note="AluSg repeat: matches 1. .309 of consensus"
repeat_region
58522. .58764
/note="MIR repeat: matches 12. .261 of consensus"
repeat_region
58904. .59119
/note="MIR repeat: matches 32. .258 of consensus"
repeat_region
59854. .60143
/note="L145 copies 2 mer aa 55% conserved"
repeat_region
63213. .63935
/note="L1M4B repeat: matches 5534. .6285 of consensus"
repeat_region
64586. .65073
/note="M111J repeat: matches 7. .516 of consensus"
repeat_region
67119. .67162
/note="L2 repeat: matches 2645. .2692 of consensus"
repeat_region
67180. .67268
/note="L1M4C repeat: matches 6608. .6696 of consensus"
repeat_region
67307. .67648
/note="L1M4C repeat: matches 6271. .6611 of consensus"

```

```

Query Match 2.0%; Score 25; DB 9; Length 144177;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1237 ATATGTTACCAATGTTGAATA 1261
Db 64051 ATATGTTACCAATGTTGAATA 64027

```

```

RESULT 27
AF020802/c 61946 bp DNA PRI 05-MAY-2001
LOCUS Homo sapiens chromosome 21 clone cosmids q98A3 and c103C0352 map
DEFINITION 21q22.2, complete sequence.
ACCESSION AF020802
VERSION AF020802.2 GI:13957611
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 61946)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S.,
Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesselmann,L., Dagand,E., Haef,T., Wehrmeyer,S.,
Bozrym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,
Reinhardt,R. and Yaspo,M.Laure.
The DNA sequence of human chromosome 21. The chromosome 21 mapping
and sequencing consortium
Nature 405 (6784), 311-319 (2000)
JOURNAL 2 (bases 1 to 61946)
MEDLINE 20289799
REFERENCE Blechschmidt,K., Nordstiek,G., Drescher,B., Weber,J., Schattevov,R.,
AUTHORS Rosenthal,A., Yaspo,M.-L., Osoegawa,K. and Soeda,E.
TITLE Direct Submission

```

JOURNAL Submitted (27-AUG-1997) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
 REFERENCE 3 (bases 1 to 61946)
 AUTHORS Blechschmidt, K., Nordstieck, G., Drescher, B., Weber, J., Schattevoy, R., Rosenthal, A., Yaspo, M.-L., Osoegawa, K. and Soeda, E.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAY-2001) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
 COMMENT On May 5, 2001 this sequence version replaced gi:2801436.
 FEATURES
 source
 1. 61946
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.2"
 /clone="cosmids Q98A3 and c103C00352"
 BASE COUNT 19958 a 12828 c 12287 g 16873 t
 ORIGIN

Query Match 1.7%; Score 22; DB 9; Length 61946;
 Best Local Similarity 100.0%; Pred. No. 5.5;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 681 GAAACCAAGAAATGACTTGGC 702
 DB 4711 GAAACCAAGAAATGACTTGGC 4690

RESULT 28
 HS1042K10 184180 bp DNA PRI 12-DEC-1999
 LOCUS
 DEFINITION Human DNA sequence from clone RP5-1042K10 on chromosome 22q13.1-13.2. Contains the ADL gene for adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ADL) and 4 novel genes (one with probable rabgap domains and Src homology domain 3). Contains ESTs, STSS, GSSS and a putative Cpg island, complete sequence.
 AL022238
 AL022238.1 GI:4176442
 HG: Adenylosuccinate lyase; ADL; Cpg Island.
 human.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 1 (bases 1 to 184180)
 Ramsey, H.
 Direct Submission
 Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Jan 22, 1999 this sequence version replaced gi:3927920.
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr22
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Inf, information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP5-1042K10 is from the library RPI-5 constructed at the Roswell Park Cancer

FEATURES

source

Institute by the group of Pieter de Jong. For further details see
 http://pacpac.med.buffalo.edu/
 VECTOR: PCVPAC2
 IMPORTANT: This sequence is not the entire insert of clone RP5-1042K10. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 The true left end of clone RP5-1042K10 is at 1 in this sequence. The true left end of clone RP4-591M18 is at 184081 in this sequence. The true right end of clone RP3-377F16 is at 39440 in this sequence. The start of this sequence overlaps with sequence 263847.

Location/Qualifiers
 1. 184180
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 /map="913.1-13.2"
 /clone="RP5-1042K10"
 /clone_1id="RPI-5"
 1. 66
 /note="Alu repeat: matches 234. .299 of consensus"
 /complement(5. .477)
 /note="match: GSS: Em:AQ152699"
 69. .380
 /note="Alu repeat: matches 1. .309 of consensus"
 765. .912
 /note="MIR repeat: matches 61. .215 of consensus"
 913. .1209
 /note="Alu repeat: matches 1. .297 of consensus"
 1225. .1402
 /note="Alu repeat: matches 140. .312 of consensus"
 1419. .1459
 /note="MIR repeat: matches 32. .71 of consensus"
 1469. .1766
 /note="Alu repeat: matches 1. .299 of consensus"
 1786. .1903
 /note="FLAM_C repeat: matches 1. .132 of consensus"
 1967. .2589
 /note="match: STRS: Em:B14103"
 /match: GSS: Em:B14103"
 14733. .14879, 15130. .15225, 17511. .17735, 18170. .18240, 23318. .23457)
 /gene="QJ1042K10.4"
 /join(<3190. .3233, 3378. .3561, 10740. .10877, 13045. .13221, 14733. .14879, 15130. .15225, 17511. .17735, 18170. .18240, 23318. .23457)
 /note="supported by GENSCAN and GENES
 match: ESTS: Em:AA615306 Em:AA065323 Em:W80019 Em:N75651 Em:AA747718 Em:AA354286"
 /codon_start=3
 /evidence-not-experimental
 /product="d11042K10.4 (novel protein)"
 /protein_id="CA18263.1"
 /db_xref="GI:4176443"
 /db_xref="SPTREMBL:095512"
 /translation="RKISOAVROODEOILAMVSALQOOQOQOQROPKHSPHPVGPKEPDLNNVPAALNGLDLOTKGIPYGGSGSGGMDYGVGKAGTSESRKOWTSMKEGLPSVATQGEAMNMGKAIIVAGKTRGSPYNOPIIIPGDIYGVGTGAGOSWLP AKSPNTKNTGSSNSAASMPPEOPGVPKRGIIQNDPEADPYVTGSGVGTATPIVD TDHQLRLNRTGSSNSLNTSLSPRAWPYASASDSTFVHSTSAKPPYKSTWSDPI GINPPLHNSKMKMKNTISSRNTPLRPPPGGLNKRSPSPMSSTPRSRGCGTODSRL ASASTWSOGSVSPSYWVYVNLHNLTPQIDGSTLRTICMGGFLTLFHLNLTGCTALIR STKQBAKAQTALH"
 3607. .3797
 /note="MIR repeat: matches 31. .226 of consensus"
 3826. .3970
 /note="L1MB8 repeat: matches 5746. .5894 of consensus"
 3885. .4574
 /note="match: GSS: Em:A0627708"
 3904. .4459
 misc_feature

```

repeat_region /note="match: GSS: Em: A0549653"
3974. .3998
/note="MER46A repeat: matches 149. .167 of consensus"
repeat_region 3999. .4074
/note="Tiger4(Zomb1) repeat: matches 1. .82 of consensus"
4581. .4879
/note="Alusg repeat: matches 1. .299 of consensus"
4986. .5106
/note="Charlie2 repeat: matches 34. .149 of consensus"
5266. .5553
/note="Alusx repeat: matches 1. .312 of consensus"
5554. .5859
/note="Charlie2 repeat: matches 2067. .2378 of consensus"
5860. .6145
/note="Alusx repeat: matches 3. .290 of consensus"
6146. .7010
/note="Charlie2 repeat: matches 2378. .3214 of consensus"
7441. .7543
/note="L2 repeat: matches 2609. .2710 of consensus"
8063. .8176
/note="57 copies 2 mer tt 64 conserved"
8239. .8547
/note="AluYa8 repeat: matches 1. .308 of consensus"
8932. .9225
/note="Alusg repeat: matches 1. .294 of consensus"
complement(9149. .9559)
/note="match: GSS: Em: A0170974"
9964. .10284
/note="L2 repeat: matches 2062. .2388 of consensus"
complement(12001. .12492)
/note="match: GSS: Em: A0793345"
12506. .13425
/note="match: GSS: Em: A0743124"
12826. .12925
/note="MIR repeat: matches 33. .134 of consensus"
13772. .14070
/note="Alusg repeat: matches 1. .298 of consensus"
complement(14139. .14674)
/note="match: GSS: Em: A0523109"
15998. .16220
/note="MIR repeat: matches 51. .262 of consensus"
18301. .18603
/note="Alu repeat: matches 1. .303 of consensus"
18627. .18860
/note="Alu repeat: matches 16. .287 of consensus"
19526. .19819
/note="Alusx repeat: matches 1. .294 of consensus"
20494. .20787
/note="Alu repeat: matches 1. .294 of consensus"
complement(20915. .21365)
/note="match: GSS: Em: A0002626"
21016. .21309
/note="Alusx repeat: matches 1. .293 of consensus"
21447. .21511
/note="Alu repeat: matches 244. .298 of consensus"
21585. .21640
/note="MER44A repeat: matches 192. .243 of consensus"
21641. .21950
/note="Alu repeat: matches 2. .310 of consensus"
21951. .22119
/note="MER44A repeat: matches 2. .192 of consensus"
22428. .22726
/note="Alusg repeat: matches 1. .299 of consensus"
complement(23111. .23679)
/note="match: GSS: Em: A0356569"
23734. .24027
/note="AluYa8 repeat: matches 1. .303 of consensus"
24423. .24690
/note="L1M1 repeat: matches 6053. .6324 of consensus"
24876. .24915
/note="MER3A repeat: matches 119. .159 of consensus"
complement(25351. .25770)
/note="match: STS: Em: AA699725"

```

```

repeat_region 25992. .26035
/note="22 copies 2 mer aa 81 conserved"
26107. .26176
/note="35 copies 2 mer gt 68 conserved"
misc_feature complement(27815. .28237)
/note="match: STS: Em: H63387"
28689. .29182
/note="match: GSS: Em: B14190"
29245. .29777
/note="match: GSS: Em: A0715590"
29258. .30130
/note="match: GSS: Em: A0738705"
misc_feature

Query Match 1.7% Score 22: DB 9: Length 184180;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1129 TTTCATATCTCACTAAGTAA 1150
|||||
Db 136023 TTTCATATCTCACTAAGTAA 136044

RESULT 29
AC084825 216425 bp DNA 03-FEB-2001
LOCUS AC084825/c Mus musculus clone RP23-114F14, WORKING DRAFT SEQUENCE, 27
DEFINITION unordered pieces.
ACCESSION AC084825 GI:12658007
KEYWORDS HTG: HTGS_PHAHEL; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 216425)
McCombie, W.R., Baker, J.P., Bahret, A., Bal, H., Balija, V.,
Dechla, N.N., de la Bastide, M., Huang, E.N., King, L., Kirchoff, K.A.,
Miller, B., Nascimento, L.U., O'Shaughnessy, A.L., Preston, R.R.,
Rodriguez, S., Santos, L., Shah, R.S., Spiegel, L.A., Toth, K., Vill, M.D.
and Zlatavern, T.
TITLE Mouse genomic Sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 216425)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
COMMENT On Feb 3, 2001 this sequence version replaced g1:11276150.
Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: GSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
Project Information
Center project name: RP23-114F14
Center clone name: RP23-114F14
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 24308: contig of 24308 bp in length
* 24309 24595: gap of unknown length
* 24596 42729: contig of 18134 bp in length
* 42730 43016: gap of unknown length
* 43017 56375: contig of 13359 bp in length
* 56376 56662: gap of unknown length

```

* 56663 69379: contig of 12717 bp in length
* 69666: gap of unknown length
* 81976: contig of 12310 bp in length
* 81977 82262: gap of unknown length
* 82263 93885: contig of 11623 bp in length
* 93886 94171: gap of unknown length
* 94172 105626: contig of 11455 bp in length
* 105627 105912: gap of unknown length
* 105913 116480: contig of 10568 bp in length
* 116481 116766: gap of unknown length
* 116767 125888: gap of 9122 bp in length
* 125889 126174: gap of unknown length
* 126175 134777: contig of 8603 bp in length
* 134778 135063: gap of unknown length
* 135064 143634: contig of 8571 bp in length
* 143635 143920: gap of unknown length
* 143921 150477: contig of 6557 bp in length
* 150478 150763: gap of unknown length
* 150764 157228: contig of 6465 bp in length
* 157229 157515: gap of unknown length
* 157516 163965: contig of 6451 bp in length
* 163966 164251: gap of unknown length
* 164252 170636: contig of 6385 bp in length
* 170637 170932: gap of unknown length
* 170933 177237: contig of 6315 bp in length
* 177238 177523: gap of unknown length
* 177524 182030: contig of 4507 bp in length
* 182031 182316: gap of unknown length
* 182317 186721: contig of 4405 bp in length
* 186722 187007: gap of unknown length
* 187008 191332: contig of 4325 bp in length
* 191333 191618: gap of unknown length
* 191619 193440: contig of 3722 bp in length
* 193441 195626: gap of unknown length
* 195627 199195: contig of 3569 bp in length
* 199196 199481: gap of unknown length
* 199482 203031: contig of 3550 bp in length
* 203032 203317: gap of unknown length
* 203318 206200: contig of 2883 bp in length
* 206201 206487: gap of unknown length
* 206488 209217: contig of 2731 bp in length
* 209218 209503: gap of unknown length
* 209504 212144: contig of 2641 bp in length
* 212145 212430: gap of unknown length
* 212431 214894: contig of 2464 bp in length
* 214895 215180: gap of unknown length
* 215181 215181: contig of 1245 bp in length.

FEATURES
source
1. 216425
/db_xref="taxon:10090"
/clone="RP23-114F14"

BASE COUNT 57862 a 49581 c 47418 g 54051 t 7513 others
ORIGIN

Query Match 1.7%; Score 22; DB 2; Length 216425;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1250 TGTTCGGAATTAATTTATT 1271
|||||
Db 60664 TGTTCGGAATTAATTTATT 60643

RESULT 30
AC073945/c 232710 bp DNA HTG 10-AUG-2001
LOCUS AC073945
DEFINITION Mus musculus clone RP23-171e18, WORKING DRAFT SEQUENCE, 16
unordered pieces.
AC073945
AC073945.18 GI:14861741
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 232710)
AUTHORS Song, L., Jiang, X. and Roe, B.A.
JOURNAL Mus musculus BAC Clone rp23-171e18
REFERENCE 2 (bases 1 to 232710)
AUTHORS Song, L., Jiang, X. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-2000) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

COMMENT
On Jul 18, 2001 this sequence version replaced g1:14647276.
NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1. 2225: contig of 2225 bp in length
* 2226 2325: gap of unknown length
* 2326 4369: contig of 2044 bp in length
* 4370 4469: gap of unknown length
* 4470 6609: contig of 2140 bp in length
* 6610 6709: gap of unknown length
* 6710 8912: contig of 2203 bp in length
* 8913 9012: gap of unknown length
* 9013 11729: contig of 2717 bp in length
* 11730 11829: gap of unknown length
* 11830 15754: contig of 3925 bp in length
* 15755 15854: gap of unknown length
* 15855 18726: contig of 2872 bp in length
* 18727 18826: gap of unknown length
* 18827 21020: contig of 2194 bp in length
* 21021 21120: gap of unknown length
* 21121 23198: contig of 2078 bp in length
* 23199 23298: gap of unknown length
* 23299 27969: contig of 4671 bp in length
* 27970 28069: gap of unknown length
* 28070 32135: contig of 4066 bp in length
* 32136 32235: gap of unknown length
* 32236 35506: contig of 3271 bp in length
* 35507 35606: gap of unknown length
* 35607 35609: contig of 2343 bp in length
* 35610 59159: gap of unknown length
* 59160 102137: contig of 42978 bp in length
* 102138 102237: gap of unknown length
* 102238 163523: contig of 61286 bp in length
* 163524 163624: gap of unknown length
* 163625 232710: contig of 69087 bp in length.

FEATURES
source
1. 232710
/db_xref="taxon:10090"
/clone="RP23-171e18"

BASE COUNT 73185 a 44003 c 43046 g 70959 t 1517 others
ORIGIN

Query Match 1.7%; Score 22; DB 2; Length 232710;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1125 CATTTTCATATCTCACTAAG 1146
|||||
Db 79460 CATTTTCATATCTCACTAAG 79439

RESULT 31

HS229041
LOCUS HS229041 323000 bp DNA PRI 22-MAY-1998
DEFINITION Homo sapiens 959 kb contig between AML1 and CBRL on chromosome
21q22: segment 1/3.
ACCESSION AJ229041 AJ229040
VERSION AJ229041.1 GI:3153101
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 323000)
Bleichschmidt, K., Rump, A., Nordsiek, G., Drescher, B., Weber, J. and
Rosenthal, A.
TITLE Sequencing and Analysis of 960 kb between AML1 and CBRL on
Chromosome 21q22.2
JOURNAL
REFERENCE 2 (bases 1 to 323000)
Bleichschmidt, K., Rump, A., Nordsiek, G., Drescher, B., Weber, J.,
Schattevoy, R., Rosenthal, A., Yaspo, M.L., Osogawa, K. and Soeda, E.
AUTHORS Direct Submission
TITLE Submitted (07-MAY-1998) Karin Bleichschmidt, Institute of Molecular
Biotechnology, Dept. of Genome Analysis, Beutenbergstrasse 11,
07745 Jena, GERMANY
COMMENT Segment entries: AJ229041, AJ229042, AJ229043
Complete contig available via ftp:
ftp.ebi.ac.uk/pub/databases/emb1/genomes/hsap1ens/contigs/aj229040/
FEATURES
source Location/Qualifiers
1. 323000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="q22"
2. 46430
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Cosmid Q98A3 (AF020802)"
160. 289
/rpt_family="FLAM_A"
/evidence=not_experimental
repeat_region complement(917. 967)
/rpt_family="MLT1B"
/evidence=not_experimental
repeat_region complement(1046. 1348)
/rpt_family="AluJo"
/evidence=not_experimental
repeat_region complement(1349. 1559)
/rpt_family="MLT1B"
/evidence=not_experimental
repeat_region complement(2055. 2418)
/rpt_family="MLT1A1"
/evidence=not_experimental
repeat_region complement(2422. 2560)
/rpt_family="MIR"
/evidence=not_experimental
repeat_region 2579. 2870
/rpt_family="AluSx"
/evidence=not_experimental
repeat_region 2878. 2964
/rpt_family="MIR"
/evidence=not_experimental
repeat_region complement(2981. 3198)
/rpt_family="MIR"
/evidence=not_experimental
exon 3702. 3821
/note="GRAIL score = 60.000%, comment = good"
/evidence=not_experimental
exon 3770. 3826
/note="GenScan, score = 9.66%, comment = Initial_exon 57
bp frame: 1 phase: 0"
/evidence=not_experimental
exon 4584. 4601

/note="GRAIL, score = 48.000%, comment = marginal"
/evidence=not_experimental
exon 4671. 4730
/note="GRAIL, score = 84.000%, comment = excellent"
/evidence=not_experimental
exon 4999. 5277
/note="GRAIL, score = 43.000%, comment = marginal"
/evidence=not_experimental
exon 5023. 5134
/note="Xpound exon prediction, score = 85% (0%)"
/evidence=not_experimental
exon complement(5730. 5823)
/note="GRAIL, score = 84.000%, comment = excellent shadow"
/evidence=not_experimental
repeat_region complement(5995. 6057)
/rpt_family="MER4_Internal"
/evidence=not_experimental
repeat_region 6560. 6647
/rpt_family="L2"
/evidence=not_experimental
exon 6719. 6898
/note="GenScan, score = 1.84%, comment = Internal_exon 180
bp frame: 1 phase: 0"
/evidence=not_experimental
repeat_region complement(8517. 8768)
/rpt_family="MLT2G"
/evidence=not_experimental
repeat_region complement(9065. 9199)
/rpt_family="MER5B8"
/evidence=not_experimental
repeat_region complement(9679. 10211)
/rpt_family="L1MD3"
/evidence=not_experimental
repeat_region complement(10721. 11087)
/rpt_family="MLT1B"
/evidence=not_experimental
repeat_region complement(11094. 11392)
/rpt_family="AluSg"
/evidence=not_experimental
repeat_region complement(11856. 12044)
/rpt_family="L2"
/evidence=not_experimental
repeat_region complement(12225. 12338)
/rpt_family="L2"
/evidence=not_experimental
exon 12294. 12390
/note="GRAIL, score = 60.000%, comment = good"
/evidence=not_experimental
repeat_region complement(14078. 14380)
/rpt_family="AluSx"
/evidence=not_experimental
exon 14490. 14717
/note="GenScan, score = 3.65%, comment = Terminal_exon 228
bp frame: 2 phase: 0"
/evidence=not_experimental
repeat_region complement(15634. 15678)
/rpt_family="L2"
/evidence=not_experimental
exon 15687. 15846
/note="GRAIL, score = 65.000%, comment = good"
/evidence=not_experimental
repeat_region complement(16160. 16454)
/rpt_family="AluSp"
/evidence=not_experimental
repeat_region complement(17435. 17595)
/rpt_family="AluJb"
/evidence=not_experimental
repeat_region complement(17657. 17936)
/rpt_family="AluSg"
/evidence=not_experimental
repeat_region complement(18646. 18944)
/rpt_family="AluSc"
/evidence=not_experimental

	ORGANISM
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (sites) Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Tauden,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy A. The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium Nature 405 (6784), 311-319 (2000)
JOURNAL	Nature 405 (6784), 311-319 (2000)
MEDLINE	20289799
REFERENCE	2 (bases 1 to 340000)
AUTHORS	Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Tauden,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Kawasaki,K., Zimmerman,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,S., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordisk,G., Hornischer,K., Baradt,P., Scharfe,M., Schoen,O., Desario,A., Reinisch,L.J., Kauert,G., Blocker,H., Ramser,J., Beck,A., Klages,S., Henning,S., Riesselmann,L., Dardand,E., Wehmaeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.
TITLE	Direct Submission
JOURNAL	Submitted (10-APR-2000). to the DBJ/EMBL/Genbank databases. The Chromosome 21 Mapping and Sequencing Consortium * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below) On May 30, 2000 this sequence version replaced gi:7717338. The chromosome 21 mapping and sequencing consortium consisting of: * RIKEN Genomic Sciences Center, Human Genome Research Group, Sagamihara 228-8555, Japan, * e.mail: hattori@gscc.riken.go.jp * URL: http://hgp.gscc.riken.go.jp/ and * Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany, * e.mail: gscj-submit@genome.imb-jena.de * URL: http://genome.imb-jena.de/ and * Keio university School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan, * e.mail: nshimizu@db.med.keio.ac.jp * URL: http://www.dmb.med.keio.ac.jp/ and * GBF, Dept. of Genome Analysis, * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de * URL: http://genome.gbf.de/ and * Max-planck institute for Molecular Genetics, Imnestrasse 73, D-14195 Berlin, Germany, * e.mail: info.chr21@molgen.mpg.de * URL: http://chr21.itz-berlin.mpg.de/ AL163269: Submitted (10-Apr-2000).
FEATURES	Location/Qualifiers 1..340000 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21"

```

/clone="p169k17, 5' partial"
/clone_id="RPC11,3-5 PAC library"
repeat_region
  /note="AluJo"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  complement(2197..2499)
  /note="AluSx"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  complement(2741..3129)
  /note="THE1C"
  /rpt_family="LTR/MaLR"
  /rpt_type=DISPERSED
  3756..3945
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  4224..4272
  /note="(GA)n"
  /rpt_family="Simple_repeat"
  /rpt_type=TANDEM
  complement(4692..5067)
  /note="THE1C"
  /rpt_family="LTR/MaLR"
  /rpt_type=DISPERSED
  5224..5728
  /note="(TA)n"
  /rpt_family="Simple_repeat"
  /rpt_type=TANDEM
  5742..6191
  /note="(TA)n"
  /rpt_family="Simple_repeat"
  /rpt_type=TANDEM
  6275..6577
  /note="AluJo"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  complement(9217..9753)
  /note="MLT2D"
  /rpt_family="LTR/Retroviral"
  /rpt_type=DISPERSED
  complement(10700..10816)
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  complement(11045..11141)
  /note="MER5B"
  /rpt_family="DNA/MER1_type"
  /rpt_type=DISPERSED
  12481..13054
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  13090..13456
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  complement(13457..13859)
  /note="L1Pa8"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  13860..13927
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  14377..14565
  /note="MER5A"
  /rpt_family="DNA/MER1_type"
  /rpt_type=DISPERSED
  complement(14684..14770)
  /note="MLT1J"
  /rpt_family="LTR/MaLR"

```

```

repeat_region
  /rpt_type=DISPERSED
  complement(14845..15169)
  /note="MLT1J"
  /rpt_family="LTR/MaLR"
  /rpt_type=DISPERSED
  17813..17917
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  18095..18315
  /note="MIR"
  /rpt_family="SINE/MIR"
  /rpt_type=DISPERSED
  18611..18708
  /note="(TA)n"
  /rpt_family="Simple_repeat"
  /rpt_type=TANDEM
  18760..18960
  /note="(TAA)n"
  /rpt_family="Simple_repeat"
  /rpt_type=TANDEM
  18775..18893
  /note="FLAM_C"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  18833..63876
  /note="Accession No. AF020803"
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="21"
  /map="21q22.1"
  /clone="Q37D2"
  /clone_id="L121NCO2-Q Cosmid library"
  18894..18960
  /note="(TAA)n"
  /rpt_family="Simple_repeat"
  /rpt_type=TANDEM
  complement(19309..19427)
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  20358..20646
  /note="L1ME3A"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  20647..20954
  /note="AluJo"
  /rpt_family="SINE/Alu"

```

Query Match 1.7%: Score 22; DB 9; Length 340000;
 Best Local Similarity 100.0%: Pred. No. 5.1;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 681 GAAACGAGAAATGAGTTTC 702
 DB 125791 GAAACGAGAAATGAGTTTC 125770

RESULT 33
 CENS061A0 958 bp DNA STS 10-JAN-2001
 LOCUS T7 end of clone AS0A019A04 of library AS0A from strain CLIB 533
 DEFINITION of Saccharomyces bayanus, sequence tagged site.
 ACCESSION A1400030
 VERSION A1400030.1 GI:12155366
 KEYWORDS STS
 SOURCE Saccharomyces bayanus
 ORGANISM Saccharomyces bayanus
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes;
 1 (bases 1 to 958)
 REFERENCE Bon,E., Neuvéglise,C., Casaregola,S., Attienave,F., Wincker,P.,
 Aigle,M. and Durrens,P.

TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 5.
JOURNAL Saccharomyces bayanus var. uvarum
PUBMED FEBS Lett. 487 (1), 37-41 (2000)
REFERENCE 11152880
AUTHORS 2 (bases 1 to 958)
Souchet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bollotin-Pukhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,D., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Maupertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Peltier,S.,
Sautin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Winkler,P. and Weissenbach,J.
TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
PUBMED 11152876
REFERENCE 3 (bases 1 to 958)
AUTHORS Genoscope.
JOURNAL Direct Submission
TITLE Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
Source
Location/Qualifiers
1..958
/organism="Saccharomyces bayanus"
/strain="Clb 533"
/variety="uvarum"
/db_xref="taxon:4931"
/clone="AS0A019A04"
/clone_1lb="AS0A0A"
/note="end : T7"
misc-feature
c23..>462
/note="similar to Saccharomyces cerevisiae ORF YBR011c [
IP1] : inorganic pyrophosphatase, cytoplasmic]
1 putative frameshift(s)"
/evidence=not-experimental
BASE COUNT 317 a 183 c 168 g 287 t 3 others
ORIGIN

Query Match 1.68; Score 21; DB 11; Length 958;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 58 GACTGGAAGTCATTCGCATT 565
|||||
58 GACTGGAAGTCATTCGCATT 78

RESULT 34
LOCUS SCF91 39525 bp DNA BCT 23-NOV-1999
DEFINITION Streptomyces coelicolor cosmid F91.
ACCESSION AL132973
VERSION 1
KEYWORDS
2-hydroxyacid-family dehydrogenase; anti-sigma factor antagonist;
binding protein dependent transport protein; catalase; catB;
Deor-family transcriptional regulator; gas vesicle synthesis
protein; glycerol-3-phosphate dehydrogenase; gvpa; gvpr; gvpg;
gvpj; gvpk; gvpl; gvpo; gvps; membrane protein; methyltransferase;
secreted cellulose-binding protein; secreted chitosanase; secreted
endo-1,4-beta-xylanase; short chain oxidoreductase; soluble-binding
lipoprotein; tellurite resistance protein; terd; tetr-family
transcriptional regulator; transmembrane efflux protein; xysA;

SOURCE zinc-containing dehydrogenase.
ORGANISM Streptomyces coelicolor A3(2).
REFERENCE 1 (bases 1 to 39525)
AUTHORS Redenbach,M., Kleser,H.M., Denapate,D., Eichner,A., Cullum,J.,
Kinaschi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
JOURNAL 97000351
MEDLINE 2 (bases 1 to 39525)
REFERENCE Oliver,K. and Harris,D.
AUTHORS Unpublished
JOURNAL 3 (bases 1 to 39525)
REFERENCE Thomson,N.R., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)
CDS are numbered using the following system eg SC787.01c. SC (S.
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.go.jp/>
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or att) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid F91 overlaps with cosmid STf15 on the AseI-F genomic
restriction fragment.

FEATURES
Source
Location/Qualifiers
1..39525
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid F91"
1..247
/gene="terd"
complement(1..228)
/note="nominal overlap with cosmid STf56"
1..247
/gene="terd"
/note="SCF91.01, terd, tellurite resistance protein,
partial CDS, len: > 81 aa. Highly similar to many
tellurite resistance proteins including: Serratia
marcescens SW:TERD_SERM(EMBL:138824) tellurium resistance
protein Terd (192 aa), fasta scores opt: 398 z-score:

Query Match	Best Local Similarity	Matches	Score	DB	Length
35	1.68;	21;	DB 1;	39525;	
AF020801/c	100.0%;	18;			
LOCUS	Conservative	0;	Mismatches	0;	Gaps 0;
DEFINITION					
AF020801					
AF020801.2					
GI:13957610					
HTG.					
human.					
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
1 (bases 1 to 44548)					
Hattori,M., Fujiyama,A., Taylor,T.D., Matanabe,H., Yada,T.,					
Park,H.-S., Toyoda,A., Ishii,K., Torokai,Y., Choi,D.-K., Soeda,E.,					
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Bleeschmidt,K.,					
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,					
Patterson,D., Rosenwald,K., Rump,A.J., Schillhabel,M.B., Schudy,A.,					
Zimmermann,W., Reichenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S.,					
Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,					
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,					
Hornischer,K., Brandt,P., Scharte,M., Schoen,O., Desario,A.,					

annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EM, EMBL; SW, SWISSPROT; TR, TREMBL; WP, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13>. Rpl1-218A18 is from the library RPl1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: PBAC3.6

IMPORTANT: This sequence is not the entire insert of clone Rpl1-218A18. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone Rpl1-270H22 is at 60060 in this sequence. The true right end of clone Rpl1-380B4 is at 100 in this sequence.

FEATURES

source location/Qualifiers
1..60159
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="Rpl1-218A18"
/clone_lib="RPl1-11.1"
2288..2579
/note="Aluub repeat: matches 1..288 of consensus"
3709..4038
/note="MLTJ repeat: matches 164..508 of consensus"
4377..4566
/note="LTR16C repeat: matches 199..387 of consensus"
4570..4689
/note="MLTJ repeat: matches 36..158 of consensus"
5281..5594
/note="Aluuo repeat: matches 5..309 of consensus"
6203..6268
/note="MLTJ repeat: matches 446..513 of consensus"
6333..6416
/note="MLTJ repeat: matches 273..354 of consensus"
6676..6832
/note="MIR repeat: matches 88..262 of consensus"
7355..7551
/note="MIR repeat: matches 47..254 of consensus"
7640..8678
/note="L1ME1 repeat: matches 5177..6158 of consensus"
8679..8972
/note="Aluuo repeat: matches 1..300 of consensus"
8973..9682
/note="L1ME1 repeat: matches 4507..5177 of consensus"
9683..9972
/note="Aluub repeat: matches 1..296 of consensus"
9973..10663
/note="L1ME1 repeat: matches 3855..4507 of consensus"
10664..11161
/note="MLTJ repeat: matches 1..503 of consensus"
11162..12453
/note="L1ME1 repeat: matches 2245..3855 of consensus"
12442..14102
/note="L1P7 repeat: matches 4477..6140 of consensus"
15771..15816
/note="L2 repeat: matches 2599..2644 of consensus"
16213..16388
/note="MER3 repeat: matches 18..189 of consensus"
17153..17217
/note="L2 repeat: matches 2644..2708 of consensus"
17232..17815
/note="L1P1 repeat: matches 5522..6154 of consensus"
18064..18115
/note="FAM repeat: matches 122..173 of consensus"
20102..20282

repeat_region
/note="L2 repeat: matches 2370..2557 of consensus"
20408..21011
/note="L1M4 repeat: matches 225..841 of consensus"
21222..21316
/note="MER3A repeat: matches 7..103 of consensus"
21369..21670
/note="Alusg repeat: matches 1..300 of consensus"
21730..21954
/note="MLTJ repeat: matches 270..494 of consensus"
21978..22097
/note="L2 repeat: matches 2589..2710 of consensus"
23357..23786
/note="L2 repeat: matches 2260..2694 of consensus"
23831..24397
/note="L1MB2 repeat: matches 4838..5421 of consensus"
24398..24692
/note="Alusg repeat: matches 1..295 of consensus"
24693..25474
/note="L1MB2 repeat: matches 5421..6168 of consensus"
26438..26696
/note="Alusg repeat: matches 36..303 of consensus"
26804..26857
/note="MER3B repeat: matches 1..54 of consensus"
26881..28220
/note="L2 repeat: matches 1267..2708 of consensus"
27519..27539
/note="Sequence from clone PCR only"
28272..28513
/note="L1MB2 repeat: matches 5803..6046 of consensus"
31634..31677
/note="L2 repeat: matches 70..238 of consensus"
32513..32678
/note="MIR repeat: matches 70..238 of consensus"
32947..33046
/note="L2 repeat: matches 2606..2707 of consensus"
33548..33636
/note="MER3A repeat: matches 97..183 of consensus"
34078..34506
/note="L2 repeat: matches 211..2750 of consensus"
34591..34654
/note="L2 repeat: matches 211..2750 of consensus"
34656..34977
/note="L2 repeat: matches 211..2750 of consensus"
35075..35251
/note="L1P1 repeat: matches 5974..6154 of consensus"
35882..36885
/note="Alusg repeat: matches 1..302 of consensus"
37331..37625
/note="Alu repeat: matches 1..300 of consensus"
37776..37964
/note="MER3A repeat: matches 7..189 of consensus"
38141..38237
/note="MIR repeat: matches 36..147 of consensus"
39030..39175
/note="HERV repeat: matches 4853..4999 of consensus"
39612..40032
/note="L1R16A repeat: matches 1..448 of consensus"
40655..41314
/note="HERV repeat: matches 4102..4787 of consensus"
41321..41470
/note="Charlie repeat: matches 2610..2761 of consensus"
41862..41905
/note="L2 repeat: matches 2599..2644 of consensus"
42378..42680
/note="Alusx repeat: matches 1..300 of consensus"
44890..45108
/note="L1ME1 repeat: matches 6099..6320 of consensus"
45109..45280
/note="FAM repeat: matches 2..175 of consensus"
45292..46245
/note="L1P3 repeat: matches 5191..6144 of consensus"
46251..46325
/note="FAM repeat: matches 52..126 of consensus"

```

repeat_region 46359..46792
                /note="LMC1 repeat: matches 5682..6107 of consensus"
repeat_region 46797..47282
                /note="LMB3 repeat: matches 5686..6182 of consensus"
repeat_region 47443..47502
                /note="LIP3 repeat: matches 5608..5666 of consensus"
repeat_region 47500..48495
                /note="LMB3 repeat: matches 4640..5695 of consensus"
repeat_region 48496..48800
                /note="Alu repeat: matches 1..302 of consensus"
repeat_region 48801..49577
                /note="LMB3 repeat: matches 3896..4640 of consensus"
repeat_region 49625..49726
                /note="Alu repeat: matches 1..102 of consensus"
repeat_region 49757..50183
                /note="MER49 repeat: matches 410..842 of consensus"
repeat_region 50769..50927
                /note="MIR repeat: matches 105..262 of consensus"
repeat_region 51097..51252
                /note="MERSA repeat: matches 7..168 of consensus"
repeat_region 52653..52701
                /note="L2 repeat: matches 2659..2710 of consensus"
repeat_region 52855..53085
                /note="LIP2 repeat: matches 5903..6155 of consensus"
repeat_region 54070..54192
                /note="L2 repeat: matches 2570..2688 of consensus"
repeat_region 54491..54580
                /note="MER30 repeat: matches 130..230 of consensus"

```

```

Query Match      1.6%; Score 21; DB 9; Length 60159;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1175 AACGAGTTCATTGGCACT 1195
Db 42189 AACGAGTTCATTGGCACT 42209

```

```

RESULT 38
LOCUS ATFB14 99461 bp DNA PLN 15-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F1B14 (GSSA
ACCESSION AL031986
VERSION AL031986
KEYWORDS GI:3805839
SOURCE thale cress
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 29277 to 29787; 37368 to 37886)
AUTHORS Bevan, M., Volckaert, G., Grymonprez, B., Voelt, M., Robben, J.,
Jesse, T., Heijnen, L., Vos, P., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99461)
AUTHORS Bevan, M., Rose, M., Hempel, S., Entian, K.-D., Jesse, T., Heijnen, L.,
Vos, P., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 99461)
AUTHORS EU Arabidopsis sequencing project.
DIRECT SUBMISSION Direct Submission
SUBMITTED (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES
Location/Qualifiers

```

```

source
1..99461
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
1..14021
/note="overlap to BAC F8D20; please refer to
EMBL:AL031135 for analysis and annotation"
16756..23438
/gene="AT4g35790"
complement(16756..16983)
/gene="AT4g35790"
exon
complement(16756..16983)
CDS
complement(16756..16983,17800..17899,17992..18015,
18128..18268,18591..18613,18688..18784,19341..19633,
19910..20110,20250..20399,20515..20663,20872..21019,
21161..21229,21319..21624,21715..21909,22187..22336,
22497..23438)
/gene="AT4g35790"
complement(16756..16983)
/note="similarity to phospholipase D, Arabidopsis
thaliana, gb:U84568
contains EST gb:AI993659.1, W43871, W43825, H36319,
N96688, N96112, W43543, Z25582, N65741, AA651235, Z26080"
/codon_start=1
/product="putative protein"
/protein_id="CA21465.1"
/db_xref="GI:3805845"
/translation="MAEKVSEDEVLHGDILIKVARRLPNDMFSEHLRLFTACN
ACAPRTDIDVDPRDGEFDKNIIRSHKVIYTSPPYVAVVPAQTLARRVKNSEOP
LMDEKFNISIAHPAYLEFOVKDDVDGAOIICTAKIPVDIASGERISGMFVLGAS
GKPKARPAFLFIOMKFPPEDOIHSYRGIGDDERGVRITPVKRGISQVRLYQAH
VMDGTDLAIGLNDKYYEHGKCWEDICATSEAHMMIYIVGMSIFPKIKVRETKR
DKMDTLELTKNSQEGVAVLLWMDKTSKDFGKITRPGMCTHDEEPKFKHSSV
ICVLSPPYASSKGLFVKQASPSSTIIMVVGILFTHKCVLTDTQANGNNKRYTA
FIVGLDLCGRYIDPEHRIIADLTVDKDFHNPTEPAGTKARQPHWDHLCNIDPA
AYVDLNEQRMKATRMKEFSLELKGKTMODALIRIGISMTLSPPKFLKDT5
IIPDDPCVWVSKEDDENHVOIFRSIDGSYKGFPEYDEAEALCEAKRLVVDK
SIORAYIOTIRSAOHFTYIENOVFLGSSVAMPYSRDAGADNLIPEMLATKIVKIRAK
ERFAYVYVPLPMPGCPKSGPVOEILLWOSOTOMOMDVATKELKAVOSAPHLDYIN
FYCLGKREQLDDMPATNGSVSDSYTFQRFMTIVHAKGMIVNDEVYLMESANNORS
MAGTKRTEIEMGAYOPNHTVAHKGRRGQVYGRMSLAEHLKTEDDEVESDLEC
LAKVNTIISENMKRFIDPKSELOGLHILKIPLOVDVGKVSPLDETFFDVGGKIIIG
AHSMALPDLTITOGKERLILRFALRTKTLISVITSVSSMAADSPORRORSRPS
PKRRARRSRSSRSRSPRLRISRSRSLPPVPSRSRSRSGRSEVENPQGT
LYVGLSTRVTDKDLAEHFAKECKIEKAVAVEGTMTVMITEVGHODATITHVMKEG
LEEGITLMEARQEDOREGLRQRCGMNVARDLDFSP"
complement(16984..17799)
/number=1
complement(17800..17899)
/gene="AT4g35790"
/number=2
complement(17900..17991)
/number=2
complement(17992..18015)
/gene="AT4g35790"
/number=3
complement(18016..18127)
/number=3
complement(18128..18268)
/gene="AT4g35790"
/number=4
complement(18269..18590)
/number=4
complement(18591..18613)
/gene="AT4g35790"
/number=5
complement(18614..18687)
intron
exon
intron
exon
intron

```

exon /number=5
 complement(18688..18784)
 /gene="AT4935790"
 intron /number=6
 complement(18785..19340)
 /number=6
 complement(19341..19633)
 /gene="AT4935790"
 intron /number=7
 complement(19634..19909)
 /number=7
 complement(19910..20110)
 /gene="AT4935790"
 intron /number=8
 complement(20111..20249)
 /number=8
 complement(20250..20399)
 /gene="AT4935790"
 intron /number=9
 complement(20400..20514)
 /number=9
 complement(20515..20663)
 /gene="AT4935790"
 exon /number=10
 complement(20664..20871)
 /number=10
 complement(20872..21019)
 /gene="AT4935790"
 intron /number=11
 complement(21020..21160)
 /number=11
 complement(21161..21229)
 /gene="AT4935790"
 exon /number=12
 complement(21230..21318)
 /number=12
 complement(21319..21624)
 /gene="AT4935790"
 intron /number=13
 complement(21625..21714)
 /number=13
 complement(21715..21909)
 /gene="AT4935790"
 exon /number=14
 complement(21910..22186)
 /number=14
 complement(22187..22336)
 /gene="AT4935790"
 intron /number=15
 complement(22337..22496)
 /number=15
 complement(22497..23438)
 /gene="AT4935790"
 exon /number=16
 complement(23439..24846,24953..25024,25138..28199,
 28285..28543,28720..29006,29101..29436,29522..29663,
 29738..29958,30041..30202,30288..30446,30524..30808,
 30897..31181,31372..31455))
 /gene="AT4935800"
 gene /number=17
 complement(24678..24846)
 /gene="AT4935800"
 exon /number=18
 complement(24678..24846)
 /gene="AT4935800"
 intron /number=19
 complement(24846,24953..25024,25138..28199,
 28285..28543,28720..29006,29101..29436,29522..29663,
 29738..29958,30041..30202,30288..30446,30524..30808,
 30897..31181,31372..31455))
 /gene="AT4935800"
 CDS /note="in contrast to gb:SL1960 gene contains an intron at
 position 29521-29437 which is present in similar proteins
 as well; the sequence differs at position 29520-29525 from
 gb:SL1960; sequence stretch has been verified on BAC

Query Match 1.68; Score 21; DB 8; Length 99461;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 788 AAACGAATGCAAAAGCAATC 808
 Db 72542 AAACGAATGCAAAAGCAATC 72562

RESULT 39
 ATT19K4
 LOCUS
 DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T19K4 (ESSAII
 Project).
 ACCESSION AL022373
 VERSION AL022373.1 GI:3036791
 KEYWORDS
 SOURCE
 ORGANISM
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 AUTHORS
 Bevan, M., Medler, H., Wambutt, R., Hohnsbein, J., Jesse, T., Heljnen, L.,
 Vos, P., Mewes, H.W., Meyer, K. and Schueller, C.
 JOURNAL
 REFERENCE
 2 (bases 1 to 106007)
 AUTHORS
 EU Arabidopsis sequencing project.
 TITLE
 JOURNAL
 Submitted (07-APR-1998) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
 'schueller@mips.biochem.mpg.de or mayer@mips.biochem.mpg.de Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bbsrc.ac.uk
 FEATURES
 source
 Location/Qualifiers
 1..106007
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 285..2210
 /gene="T19K4.10"
 285..374
 /gene="T19K4.10"
 /number=1
 join(285..374,483..718,803..1027,1117..1252,1351..1412,
 1870..1904,1987..2210)
 /gene="T19K4.10"
 /note="strong similarity to predicted protein A_TW017A05.8
 of Arabidopsis thaliana BAC TW017A05, Patchx:G2435517

Contains eukaryotic and viral aspartyl proteases active sites [VALDTGSDLFW] and [ALFDGTGSEFTYL]"

/codon_start=1
/product="putative protein (fragment)"
/protein_id="CA18482.1"
/db_xref="GI:3036792"
/db_xref="SPTREMBL:O65625"
/translation="MVALDTGSDLFWPCDCGCAPEGATYASEFELSIYKPVSTT
NKVTYNSLCAQNRNOCIGFTSCPYMYSAQSTSGILMEDVMHLTDEKPERV
EAYVFGGVOYSGSFLDIAPNGLFGEMKISPSYLAREGLVADSFSCFGHDV
GRISFGDKSSDOETEPNLSHPNNYITRYRVGFTLLIDDEFTALFDGTSFTYL
VDPMTTVSESAQDKRHSPOSRIPEFCYDMREKLVIAKKFDCTDIEETVTAGTN
KTAAVAPMAAGIKTHNNSSELKNTITSKSSNPQISKTVDWSEFRFTLLPL
V"

375..482
/gene="T19K4.10"
/number=1
483..718
/gene="T19K4.10"
/number=2
719..802
/gene="T19K4.10"
/number=2
803..1027
/gene="T19K4.10"
/number=3
1028..1116
/gene="T19K4.10"
/number=3
1117..1252
/gene="T19K4.10"
/number=4
1253..1350
/gene="T19K4.10"
/number=4
1351..1412
/gene="T19K4.10"
/number=5
1413..1869
/gene="T19K4.10"
/number=5
1870..1904
/gene="T19K4.10"
/number=6
1905..1986
/gene="T19K4.10"
/number=6
1987..2210
/gene="T19K4.10"
/number=7
3922..6899
/gene="T19K4.20"
3922..4443
/gene="T19K4.20"
/number=1
join(3922..4443,4917..5406,5739..5809,6321..6392,
6636..6899)
/gene="T19K4.20"
/note="similarity to protein coded for by gene T12P5.5 in
Caenorhabditis elegans cosmid T12P5, Patchx:G2773225
contains EST gb:z26699"
/codon_start=1
/product="putative protein"
/protein_id="CA18483.1"
/db_xref="GI:3036793"
/db_xref="SPTREMBL:O65626"
/translation="MASATSNPPASSSSSPRISGNCSPTASVAQSPRRRSROYSSP
WTQIVRGSEPTIAAAVAPSSPSQSRAPIEPIASVVAAPTAALVIEAAGDEKSE
ASGGDNAGKPKVVRKPSNGASEVPGVAGASSWPALSETTAPSKSSSDSLKIGDV
PSSSSASSSVLTGIIANAVAPAKOAGRANPTPNHSRSPFORGASGSANGTV
SOPSAOGSEFTLPSHNPSPGCONKNGFASONGHGTENPSORDSYRNONGHSHG
RRNEHGNOMTFFORSPNGREGNAOSGCTPAFYRHPSPYOPIDOPMAADPPSHIP
FPDILYNGPMNNEGFVPLRVYAGFKKVAELTDNIQQLVEALONSFHEVCGDFTIRK

DMQNVLRNPPTGSGPOSVRADAVAKRLGNLSVDSSADPIGSSSSQLOPTEALSD
DOOOSSTAPVSNHNPADGANR"

4444..4916
/gene="T19K4.20"
/number=1
4917..5406
/gene="T19K4.20"
/number=2
5407..5738
/gene="T19K4.20"
/number=2
5739..5809
/gene="T19K4.20"
/number=3
5810..6320
/gene="T19K4.20"
/number=3
6321..6392
/gene="T19K4.20"
/number=4
6393..6635
/gene="T19K4.20"
/number=4
6636..6899
/gene="T19K4.20"
/number=5
11235..11969
/gene="T19K4.30"
11235..11969
/gene="T19K4.30"
/note="weak similarity to GBF4, Arabidopsis thaliana,
PAX:G403418
contains bZIP transcription factors basic domain signature
[KRMIKRESAPRSRRAR]"

/codon_start=1
/product="putative bZIP transcription factor"
/protein_id="CA18484.1"
/db_xref="GI:3036794"
/db_xref="SPTREMBL:O65627"
/translation="MLSSAKHQRHRLSATNKNOTLRKVSISSSSPSSSSSSSTSS
SSPLPSODSQAKRSILVMEVMDINLASIHILNHRSPHPOHNEPFRQONHNGN
PNSIFQDFLKSLNOEPAPTSQTSAPNGSSTVTVLSSPPRPATVLSLNGAGF
EFLNODPIVTSNSNLHTHHLSNAHFNFSFELVLPSSSGKKRGQDSNGSGNRHH
KRMIKRESAPRSRRARKEOCASPLPS"
/number=1
11235..11969
/gene="T19K4.30"
/number=1
14550..17503
/gene="T19K4.40"
complement(14550..14672)
/gene="T19K4.40"
/number=1
complement(join(14550..14672,14834..14891,15525..15919,
16020..16158,16200..16321,16597..16683,16777..17140,
17208..17503))
/gene="T19K4.40"
complement(join(14550..14672,14834..14891,15525..15919,
16020..16158,16200..16321,16597..16683,16777..17140,
17208..17503))
/gene="T19K4.40"
/note="similarity to predicted protein P29F11.3,
Caenorhabditis elegans, Patchx:E248390
contains EST gb:T21369"
/codon_start=1
/product="putative protein"
/protein_id="CA18485.1"
/db_xref="GI:3036795"
/db_xref="SPTREMBL:O65628"
/translation="MACNSSGCEGCGCDREKNGSKITVDVSGGNGHEESVCKCKN
APMTFGGGEDEGRCADCEFRNNVFGFRLAVTSHAMITPSDNVAFSGSSSSSKYD
VRVSLQVNEHQLIALKNYEASRRSLPVGVAFAVDETAAPALSTEMIDALEVR
VTVSCLSPKADLHVVESEIFGSDSIDARDRLKLDVDPDQGRKEDLLHKMLSL
OKVAENGINRLVLGSECTRIASHIVLTRYKGRYSLSDQIHDAHWKRYIVLPLD

CVRLETRCLDGEPEYSCITIOILSVYSGGRKLTVELACRSCGGINLVSSFVAL
 LQENPSRECTIVRTAKLTPFYNKIPEDSDNPMATDRKRNKLDKDGSMTEA
 FCPICNPINRSDSSELDPEEGESDVLVYACSSCRPILRODSCSFPLPD
 HMTSOVKHOKVDQOAYLRHYNTRKLMFEDPITKSDGMGRAPFLKSKSTIAQRSLIP
 ETGKKKXVRELIKALEM"

Intnon complement(14673.14833)
 /number=1
 complement(14834.14891)
 /gene="T19K4.40"

exon /number=2
 complement(14892.15524)
 /number=2
 complement(15525.15919)
 /gene="T19K4.40"

Intnon
 exon

Query Match 1.6% Score 21: DB 8: Length 106007;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 788 AAAACGATGCGAAGGATC 808
 ||||||||||||||||||
 Db 15469 AAAACGATGCGAAGGATC 15489

RESULT 40
 AC008034/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AC008034 120134 bp DNA HTG 17-JUL-2001
 Homo sapiens chromosome 3 clone CTB-134A23, WORKING DRAFT SEQUENCE,
 3 unordered pieces.
 AC008034
 AC008034.28 GI:14787124
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 120134)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alstbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbata,J.,
 Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C.,
 Burck,P., Burkelt,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyte,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C.,
 Hollins,B., Homsl,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
 Lounsged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nockenkwo,S.,
 Oguh,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojudoan,I., Rolfe,M.,
 Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
 Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

Washington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G., and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 120134)
 Morley,K.C.
 Direct Submission
 Submitted (13-JUL-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 17, 2001 this sequence version replaced gi:14547711.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HMHO
 Center clone name: CTB-134A23
 ----- Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-terminator Big Dye; 30% of reads
 Chemistry: Dye-terminator Big Dye; 70% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 121889 bases at least 940
 Consensus quality: 124362 bases at least 940
 Consensus quality: 125920 bases at least 920
 Estimated insert size: 121022; sum-of-confids estimation
 Quality coverage: 10.7x in Q20 bases; sum-of-confids estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 77264: contig of 77264 bp in length
 * 77265 77364: gap of unknown length
 * 77365 117203: contig of 39839 bp in length
 * 117204 117303: gap of unknown length
 * 117304 120134: contig of 2831 bp in length.
 * Location/Qualifiers
 1. 120134
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CTB-134A23"
 /chromosome="13"
 BASE COUNT 31243 a 29435 c 29477 g 29748 t 231 others
 ORIGIN

Query Match 1.6% Score 21: DB 2: Length 120134;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CTGTGCGACGCGCGCGCAG 76
 ||||||||||||||||||
 Db 99470 CTGTGCGACGCGCGCGCAG 99450

RESULT 41
 AC011414/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS

AC011414 123224 bp DNA PRI 20-JUL-2001
 Homo sapiens chromosome 5 clone CTB-83D3, complete sequence.
 AC011414
 AC011414.4 GI:14971175
 HTG.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia: Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia: Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 123224)
JOURNAL Direct Submission
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 123224)
TITLE Direct Submission
JOURNAL Direct Submission
REFERENCE Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL 3 (bases 1 to 123224)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
JOURNAL Submitted (20-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 20, 2001 this sequence version replaced gi:13677003.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.3.
STS Content:
WI-14133 G22788.
Location/Qualifiers
1. 123224
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-83D3"
BASE COUNT 43012 a 23877 c 23705 g 32630 t
ORIGIN

Query Match 1.6%; Score 21; DB 9; Length 123224;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 AAAACGAGAAATGAGTTTC 702
|||||
Db 109559 AAAACGAGAAATGAGTTTC 109539

RESULT 42
AC008491/c 143661 bp DNA PRI 31-MAY-2001
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTC-426L22, complete sequence.
AC008491
AC008491.6 GI:14269675
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 143661)
JOURNAL Direct Submission
REFERENCE Direct Submission
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Unpublished
TITLE 2 (bases 1 to 143661)
JOURNAL Direct Submission
REFERENCE Direct Submission
AUTHORS Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 143661)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
JOURNAL Submitted (31-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On May 31, 2001 this sequence version replaced gi:13676952.
Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.3.
Location/Qualifiers
1. 143661
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-426L22"
BASE COUNT 45816 a 29494 c 29282 g 39069 t
ORIGIN

Query Match 1.6%; Score 21; DB 9; Length 143661;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 AAAACGAGAAATGAGTTTC 702
|||||
Db 32913 AAAACGAGAAATGAGTTTC 32893

RESULT 43
HS179115A 146810 bp DNA PRI 22-NOV-1999
LOCUS
DEFINITION Human DNA sequence from PAC 179115, BRCA2 gene region chromosome.
13612-q13 contains Kiotho ESTs and Cpg island.
292540 292541
292540.1 GI:2370068
13612-q13; Cpg island.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 146810)
TITLE Direct Submission
JOURNAL Submitted (20-AUG-1997) Chromosome 13 Project Group
(http://www.sanger.ac.uk/HGP/Chr13/) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humuery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 14, 2001 this sequence version replaced gi:1869765
gi:1869764.
COMMENT IMPORTANT: This sequence is not the entire insert of clone 179115.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
This sequence was generated from part of bacterial clone contigs of
human chromosome 13, constructed by the Sanger Centre chromosome 13
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 179115 is at 1 in this sequence. The
true right end of clone 49010 is at 54668.
The true left end of clone 46123 is at 146707.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
179115 is from the library RPL1 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/
FEATURES
source 1. 146810
Location/Qualifiers


```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-q13"
/clone="XX-179115"
/clone_lib="RPC11"
1029. .1334
/note="AluJo repeat: matches 296. .2 of consensus"
1778. .1945
/note="3 copies of 56 mer 85 & conserved"
3350. .3496
/note="AluX repeat: matches 302. .156 of consensus
incomplete repeat"
3606. .3772
/note="MLT2_internal repeat: matches 4296. .4125 of
consensus"
3808. .4594
/note="MLT2_internal repeat: matches 4041. .3239 of
consensus"
4627. .5080
/note="MLT2_internal repeat: matches 3146. .2693 of
consensus"
5238. .5700
/note="MLT2_internal repeat: matches 2425. .1983 of
consensus"
5727. .5897
/note="MER42B repeat: matches 1109. .1300 of consensus"
6013. .6042
/note="15 copies of 2 mer 93 & conserved"
7140. .7255
/note="49J10 data only"
8271. .8292
/note="11 copies of 2 mer 100 & conserved"
9168. .9279
/note="2 copies of 56 mer 84 & conserved"
9310. .9576
/note="AluJo repeat: matches 257. .1 of consensus
incomplete repeat"
9910. .9949
/note="10 copies of 4 mer 85 & conserved"
10283. .10356
/note="MLT1C repeat: matches 1. .86 of consensus"
10354. .11007
/note="LIMA6 repeat: matches 403. .1042 of consensus"
11022. .11165
/note="MLT1C repeat: matches 67. .203 of consensus"
11022. .11364
/note="MLT1B repeat: matches 67. .390 of consensus"
11681. .11793
/note="49J10 data only"
11924. .12175
/note="MIR repeat: matches 259. .11 of consensus"
12395. .12604
/note="MIR repeat: matches 22. .218 of consensus"
13477. .13774
/note="AluX repeat: matches 1. .298 of consensus"
13879. .14366
/note="MLT1D repeat: matches 500. .2 of consensus"
15236. .15593
/note="MLT1A repeat: matches 365. .1 of consensus"
16081. .16179
/note="MIR repeat: matches 133. .35 of consensus"
17340. .17639
/note="AluJo repeat: matches 302. .1 of consensus"
20484. .20783
/note="AluSq repeat: matches 302. .1 of consensus"
21616. .21916
/note="AluSq repeat: matches 1. .301 of consensus"
22823. .23227
/note="MLT1G repeat: matches 512. .235 of consensus"
22923. .23026
/note="MLT1F repeat: matches 541. .439 of consensus"
23230. .23358

/note="MLT1G repeat: matches 160. .31 of consensus"
25654. .25859
/note="LIME3A repeat: matches 585. .366 of consensus"
26810. .26857
/note="24 copies of 2 mer 81 & conserved"
26939. .27058
/note="AluJo repeat: matches 130. .1 of consensus
incomplete repeat"
28240. .28439
/note="MIR repeat: matches 223. .20 of consensus"
28600. .28699
/note="MIR repeat: matches 261. .159 of consensus"
28700. .29065
/note="MLT1A2 repeat: matches 374. .1 of consensus"
29067. .29208
/note="MIR repeat: matches 162. .15 of consensus"
30635. .30919
/note="MER34 repeat: matches 532. .235 of consensus"
30925. .31199
/note="MER34 repeat: matches 267. .8 of consensus"
31050. .31118
/note="49J10 data only"
31268. .31344
/note="MIR2 repeat: matches 70. .146 of consensus"
32875. .33352
/note="49J10 data only"
34039. .34333
/note="AluJo repeat: matches 293. .1 of consensus"
37035. .37341
/note="AluJo repeat: matches 5. .302 of consensus"
38430. .38586
/note="MIR repeat: matches 88. .262 of consensus"
39109. .39270
/note="MIR repeat: matches 47. .207 of consensus"
39384. .39740
/note="LIME1 repeat: matches 915. .540 of consensus"
39536. .40231
/note="LIME2 repeat: matches 744. .2 of consensus"
40086. .40364
/note="L1 repeat: matches 5390. .5107 of consensus"
40435. .40728
/note="AluJo repeat: matches 1. .302 of consensus"
40770. .41372
/note="L1 repeat: matches 5013. .4394 of consensus"
41439. .41728
/note="AluJo repeat: matches 296. .1 of consensus"
41740. .42353
/note="L1 repeat: matches 4313. .3694 of consensus"
42420. .42917
/note="MLT1D repeat: matches 1. .503 of consensus"
44193. .45080
/note="L1P2 repeat: matches 893. .1 of consensus"
44935. .45858
/note="L1 repeat: matches 5390. .4467 of consensus"
44988. .45003
/note="49J10 data only"
48909. .48964
/note="MIR2 repeat: matches 144. .88 of consensus"
48968. .49571
/note="L1P1 repeat: matches 901. .269 of consensus"
49815. .49871
/note="7SLRNA repeat: matches 258. .314 of consensus"
50360. .50553
/note="MIR repeat: matches 18. .228 of consensus"
52978. .53072
/note="MER5A repeat: matches 103. .7 of consensus"
53125. .53426
/note="AluSq repeat: matches 1. .300 of consensus"
53734. .53853
/note="MIR2 repeat: matches 25. .146 of consensus"
54190. .54318
/note="49J10 data only"
55591. .56086
```

repeat_region /note="L1 repeat: matches 4844; .5349 of consensus"
55984.56140
/note="L1MB5 repeat: matches 5.157 of consensus"
56141.56434
repeat_region /note="L1MB4 repeat: matches 1.295 of consensus"
56436.57216
repeat_region /note="L1MB4 repeat: matches 148.932 of consensus"

Query Match 1.6%; Score 21; DB 9; Length 146810;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1175 AAGCAGTGTGATTCGAGCT 1195
|||||
DB 73954 AAGCAGTGTGATTCGAGCT 73974

RESULT 44
AC068315 155313 bp DNA HTG 27-SEP-2000
LOCUS Homo sapiens chromosome 3 clone RP11-224P21 map 3p, WORKING DRAFT
DEFINITION SEQUENCE, 28 unordered pieces.
AC068315
VERSION AC068315.4 GI:10312236
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT;
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 155313)
AUTHORS Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S.,
Huang, F., Jin, Y., Kang, N., Li, C., Li, G., Li, J., Li, L.,
Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, Y.,
Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y.,
Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R.,
Yan, C., Yang, X., Yu, B., Zeng, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y.,
Zhang, L., Zhang, M., Zhang, X., Zhang, G., Zhang, H., Zhang, H.,
Zhang, Z., Zhu, B., Yu, J. and Yang, H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 155313)
AUTHORS Bao, W., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,
Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y.,
Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y.,
Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D.,
Huang, F., Zhang, G., Li, T., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J.
and Yang, H.
Direct Submission
Submitted (02-MAY-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
On Sep 27, 2000 this sequence version replaced gi:1011156.
Genome Center
Center: Beijing Center
Center code: Beijing
Website: http://hgsc.igmp.ac.cn
http://www.genomics.org.cn
Contact: hgsc@igmp.ac.cn
Project Information
Center project name: RP11-224P21
Center clone name: RP11-224P21
Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 55% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 154843 bases at least Q40
Consensus quality: 165011 bases at least Q30
Consensus quality: 168985 bases at least Q20
Insert size: 139544; sum-of-contigs
Quality coverage: 5.36x in Q20 bases; sum-of-contigs

***** NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available, and the accession number will
* be preserved.

1 1139: contig of 1139 bp in length
1140 1239: gap of unknown length
1240 2448: contig of 1209 bp in length
2449 2548: gap of unknown length
2549 3935: contig of 1387 bp in length
3936 4035: gap of unknown length
4036 6131: contig of 2096 bp in length
6132 6231: gap of unknown length
6232 7429: contig of 1198 bp in length
7430 7529: gap of unknown length
7530 9054: contig of 1525 bp in length
9055 9154: gap of unknown length
9155 10950: contig of 1796 bp in length
10951 11050: gap of unknown length
11051 12638: contig of 1587 bp in length
12639 12737: gap of unknown length
12738 15121: contig of 2384 bp in length
15122 15221: gap of unknown length
15222 17320: contig of 1999 bp in length
17321 20436: contig of 3116 bp in length
20437 20536: gap of unknown length
20537 22761: contig of 3225 bp in length
22762 23762: gap of unknown length
23763 26190: contig of 2329 bp in length
26191 26290: gap of unknown length
26291 28871: contig of 2581 bp in length
28872 33062: gap of unknown length
33063 33162: contig of 4091 bp in length
33163 37033: gap of unknown length
37034 37133: contig of 3671 bp in length
37134 43465: gap of unknown length
43466 43565: contig of 6332 bp in length
43566 48805: gap of unknown length
48806 48905: contig of 5240 bp in length
48906 48905: gap of unknown length
48907 48905: gap of unknown length
48908 48905: gap of unknown length
48909 48905: gap of unknown length
48910 48905: gap of unknown length
48911 48905: gap of unknown length
48912 48905: gap of unknown length
48913 48905: gap of unknown length
48914 48905: gap of unknown length
48915 48905: gap of unknown length
48916 48905: gap of unknown length
48917 48905: gap of unknown length
48918 48905: gap of unknown length
48919 48905: gap of unknown length
48920 48905: gap of unknown length
48921 48905: gap of unknown length
48922 48905: gap of unknown length
48923 48905: gap of unknown length
48924 48905: gap of unknown length
48925 48905: gap of unknown length
48926 48905: gap of unknown length
48927 48905: gap of unknown length
48928 48905: gap of unknown length
48929 48905: gap of unknown length
48930 48905: gap of unknown length
48931 48905: gap of unknown length
48932 48905: gap of unknown length
48933 48905: gap of unknown length
48934 48905: gap of unknown length
48935 48905: gap of unknown length
48936 48905: gap of unknown length
48937 48905: gap of unknown length
48938 48905: gap of unknown length
48939 48905: gap of unknown length
48940 48905: gap of unknown length
48941 48905: gap of unknown length
48942 48905: gap of unknown length
48943 48905: gap of unknown length
48944 48905: gap of unknown length
48945 48905: gap of unknown length
48946 48905: gap of unknown length
48947 48905: gap of unknown length
48948 48905: gap of unknown length
48949 48905: gap of unknown length
48950 48905: gap of unknown length
48951 48905: gap of unknown length
48952 48905: gap of unknown length
48953 48905: gap of unknown length
48954 48905: gap of unknown length
48955 48905: gap of unknown length
48956 48905: gap of unknown length
48957 48905: gap of unknown length
48958 48905: gap of unknown length
48959 48905: gap of unknown length
48960 48905: gap of unknown length
48961 48905: gap of unknown length
48962 48905: gap of unknown length
48963 48905: gap of unknown length
48964 48905: gap of unknown length
48965 48905: gap of unknown length
48966 48905: gap of unknown length
48967 48905: gap of unknown length
48968 48905: gap of unknown length
48969 48905: gap of unknown length
48970 48905: gap of unknown length
48971 48905: gap of unknown length
48972 48905: gap of unknown length
48973 48905: gap of unknown length
48974 48905: gap of unknown length
48975 48905: gap of unknown length
48976 48905: gap of unknown length
48977 48905: gap of unknown length
48978 48905: gap of unknown length
48979 48905: gap of unknown length
48980 48905: gap of unknown length
48981 48905: gap of unknown length
48982 48905: gap of unknown length
48983 48905: gap of unknown length
48984 48905: gap of unknown length
48985 48905: gap of unknown length
48986 48905: gap of unknown length
48987 48905: gap of unknown length
48988 48905: gap of unknown length
48989 48905: gap of unknown length
48990 48905: gap of unknown length
48991 48905: gap of unknown length
48992 48905: gap of unknown length
48993 48905: gap of unknown length
48994 48905: gap of unknown length
48995 48905: gap of unknown length
48996 48905: gap of unknown length
48997 48905: gap of unknown length
48998 48905: gap of unknown length
48999 48905: gap of unknown length
49000 48905: gap of unknown length

FEATURES

source

1..155313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-224P21"
1..1139
/note="assembly_name:Contig32"

misc_feature 1240..2448
/note="assembly_name:Contig33"
misc_feature 2549..3935
/note="assembly_name:Contig34"
misc_feature 4036..6131
/note="assembly_name:Contig35"
misc_feature 6232..7429
/note="assembly_name:Contig36"
misc_feature 7530..9054
/note="assembly_name:Contig37"
misc_feature 9155..10950
/note="assembly_name:Contig38"
misc_feature 11051..12637
/note="assembly_name:Contig39"
misc_feature 12738..15121
/note="assembly_name:Contig40"
misc_feature 15222..17220
/note="assembly_name:Contig41"
misc_feature 17321..20436
/note="assembly_name:Contig42"
misc_feature 20537..23761
/note="assembly_name:Contig43"
misc_feature 23862..26190
/note="assembly_name:Contig44"
misc_feature 26291..28871
/note="assembly_name:Contig45"
misc_feature 28972..33062
/note="assembly_name:Contig46"
misc_feature 33163..37033
/note="assembly_name:Contig47"
misc_feature 37134..43465
/note="assembly_name:Contig48"
misc_feature 43566..48805
/note="assembly_name:Contig49"
misc_feature 48906..56158
/note="assembly_name:Contig50"
misc_feature 56259..63409
/note="assembly_name:Contig51"
misc_feature 63510..72539
/note="assembly_name:Contig52"
clone_end:SP6
vector_side:right"
misc_feature 72640..83389
/note="assembly_name:Contig53"
misc_feature 83490..91426
/note="assembly_name:Contig54"
misc_feature 91527..99509
/note="assembly_name:Contig55"
misc_feature 99610..108238
/note="assembly_name:Contig56"
misc_feature 108339..119079
/note="assembly_name:Contig57"
misc_feature 119180..135925
/note="assembly_name:Contig58"
misc_feature 136026..155313
/note="assembly_name:Contig59"
BASE COUNT 37836 a 38371 c 38396 g 38005 t 2705 others
ORIGIN

Query Match 1.6% Score 21 DB 2 Length 155313;
Best Local Similarity 100.0% Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CTGTGCGAGCGCGCGCGAG 76
|||||
Db 127763 CTGTGCGAGCGCGCGCGAG 127783

RESULT 45
AC066599 157007 bp DNA HTG 19-SEP-2000
LOCUS AC066599 Homo sapiens chromosome 3 clone RP11-596J9 map 3p, WORKING DRAFT
DEFINITION

SEQUENCE, 14 unordered pieces.
AC066599
AC066599.3 GI:10190763
HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 157007)
REFERENCE
AUTHORS
Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S., Huang, F., Jia, Y., Kang, N., Li, C., Li, C., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Yu, J., and Yang, H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 157007)
Wang, J., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J., and Yang, H.
Direct Submission
Submitted (25-APR-2000) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R. China
On Sep 19, 2000 this sequence version replaced gi:8101249.
COMMENT
-----Genome Center
Center: Beijing Center
Center code: Beijing
Website: http://hgsc.igtp.ac.cn
http://www.genomics.org.cn
Contact: hgsc@igtp.ac.cn
-----Project Information
Center project name: RP11-596J9
Center clone name: RP11-596J9
-----Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator Big Dye; 45% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 150014 bases at least Q40
Consensus quality: 153515 bases at least Q30
Consensus quality: 155718 bases at least Q20
Insert size: 148003; sum-of-contigs
Quality coverage: 4.98x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1641: contig of 1641 bp in length
* 1642 1741: gap of unknown length
* 1742 5811: contig of 4070 bp in length
* 5812 5911: gap of unknown length
* 5912 7862: contig of 1951 bp in length
* 7863 7962: gap of unknown length
* 7963 11614: contig of 3652 bp in length
* 11615 11714: gap of unknown length
* 11715 15628: contig of 3914 bp in length
* 15629 15728: gap of unknown length
* 15729 22199: contig of 6471 bp in length
* 22200 22299: gap of unknown length

```

* 22300 32786: contig of 10487 bp in length
* 32787 32886: gap of unknown length
* 32887 39504: contig of 6618 bp in length
* 39505 39605: gap of unknown length
* 39605 48115: contig of 8511 bp in length
* 48115 48216: gap of unknown length
* 48216 57789: contig of 9573 bp in length
* 57789 57888: gap of unknown length
* 57888 69171: contig of 11283 bp in length
* 69171 69272: gap of unknown length
* 69272 86305: contig of 17034 bp in length
* 86305 86406: gap of unknown length
* 86406 117887: contig of 31482 bp in length
* 117887 117988: gap of unknown length
* 117988 157007: contig of 39020 bp in length.

```

FEATURES

Source

```

1. 157007
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="3"
  /map="3p"
  /clone="RP11:59609"

```

```

misc_feature 1. 1641
  /note="assembly_name:Contig5"
misc_feature 1742. 5811
  /note="assembly_name:Contig6"
misc_feature 5912. 7862
  /note="assembly_name:Contig7"
misc_feature 7963. 11614
  /note="assembly_name:Contig8"
misc_feature 11715. 15628
  /note="assembly_name:Contig9"
misc_feature 15729. 22199
  /note="assembly_name:Contig10"
misc_feature 22300. 32786
  /note="assembly_name:Contig11"
misc_feature 32887. 39504
  /note="assembly_name:Contig12"
misc_feature 39605. 48115
  /note="assembly_name:Contig13"
misc_feature 48216. 57788
  /note="assembly_name:Contig14"
misc_feature 57889. 69171
  /note="assembly_name:Contig15"
misc_feature 69272. 86305
  /note="assembly_name:Contig16"
misc_feature 86406. 117887
  /note="assembly_name:Contig17"
misc_feature 117988. 157007
  /note="assembly_name:Contig18"

```

```

BASE COUNT 39725 a 38767 c 38530 g 38682 t 1303 others
ORIGIN

```

```

Query Match 1.6%; Score 21; DB 2; length 157007;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 56 CTGTGGCAGCGCGCGCGCAG 76
   |||
Db 82733 CTGTGGCAGCGCGCGCGCAG 82753

```

```

Search completed: March 29, 2002, 21:43:28
Job time: 6623 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 20:20:40 ; Search time 171.98 Seconds

(Without alignments)
6355.912 Million cell updates/sec

Title: US-09-415-540-2

Perfect score: 1275
Sequence: 1 CAACAGCTTNGGCGCTCTCT.....GAAATAAATATTTGCTG 1275

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

N_Geneseq_1101:*

1: /SIDSI/gcgdata/geneseq/geneseqn/NA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseqn/NA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseqn/NA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseqn/NA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseqn/NA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseqn/NA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseqn/NA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseqn/NA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseqn/NA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseqn/NA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseqn/NA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseqn/NA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseqn/NA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseqn/NA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseqn/NA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseqn/NA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1275	100.0	1275	20	AAZ23249 Human pyrophosphatase
2	1275	100.0	1275	20	AAV63542 CDNA encoding a hu
3	1275	100.0	1277	22	AA508596 DNA encoding human
4	1214	95.2	1351	20	AA500685 Human secreted pro
5	1153	90.4	1350	20	AA500716 Human secreted pro
6	1055	82.7	1200	20	AA233425 Human prostate can
7	788	61.8	856	20	AA277501 Human ovarian tumo
8	659	51.7	660	21	AA106628 Human secreted pro
9	415	32.5	578	21	AA78135 CDNA encoding huma
10	415	32.5	578	22	AA128873 Colon tumour relat
11	317	24.9	416	16	AA21836 Human gene signatu

12	202	15.8	357	20	AAZ77472 Human ovarian tumo
13	121	9.5	121	21	AAAC14929 Human secreted pro
14	68	5.3	102	21	AAA45576 Human secreted exp
15	66	5.2	185	22	AA122766 Probe #12699 for g
16	66	5.2	185	22	AA148068 Probe #16754 used
17	66	5.2	186	22	AA108439 Probe #8430 used t
18	66	5.2	411	22	AA113542 Probe #3475 for ge
19	66	5.2	411	22	AA134904 Probe #3590 used t
20	66	5.2	411	22	AA103431 Probe #3422 used t
21	42	3.3	411	22	AAH35811 Human colon cancer
22	19	1.5	1198	20	AAH89287 Human regulatory p
23	23	1.5	1202	20	AAZ52867 Human prostate tum
24	19	1.5	1215	21	AAAC98069 Human colon cancer
25	19	1.5	1536	22	AA159678 Human polynucleoti
26	19	1.5	1536	22	AA159679 Human polynucleoti
27	19	1.5	2264	20	AAV81267 Bcl-2 interaction
28	19	1.5	2528	21	AAZ51757 Full length CDNA s
29	19	1.5	2534	21	AAZ51804 Full length expand
30	19	1.5	3636	20	AAV84234 Rat (Na,K)-ATPase
31	19	1.5	3636	20	AAV84235 Rat (Na,K)-ATPase
32	19	1.5	3636	20	AAV84236 Rat (Na,K)-ATPase
33	19	1.5	3636	20	AAV84237 Rat (Na,K)-ATPase
34	19	1.5	7328	22	AAZ44683 Novel protein kinase
35	18	1.4	30	21	AAZ38953 Expression vector
36	18	1.4	346	20	AAV90298 EST clone D1482.
37	18	1.4	402	21	AAAC01550 Human secreted pro
38	18	1.4	543	21	AAE11454 Aspergillus niger
39	18	1.4	582	22	AAH13155 Human cDNA clone (
40	18	1.4	582	22	AAH67523 Aspergillus niger
41	18	1.4	593	21	AAE11560 VEGF-B16 coding s
42	18	1.4	624	17	AAZ37915 CDNA clone encodin
43	18	1.4	624	20	AAV63569 Human VEGF-3 codin
44	18	1.4	624	22	AAA01005 Pinus radiata flav
45	18	1.4	762	21	AAH68125

ALIGNMENTS

RESULT	1
ID	AAZ23249 standard; CDNA; 1275 BP.
XX	AAZ23249:
AC	24-JAN-2000 (first entry)
XX	
DE	Human pyrophosphatase (HPP) encoding CDNA.
XX	
KW	Human; pyrophosphatase; HPP; genetic engineering; inflammatory disease;
KW	cancer; diagnostic assay; ss.
XX	
OS	Homo sapiens.
XX	
PN	US5981232-A.
XX	
PD	09-NOV-1999.
XX	
FE	13-AUG-1998; 9805-0134593.
XX	
PR	31-OCT-1996; 96US-0741437.
XX	
PA	(INCT-) INCTE PHARM INC.
XX	
PI	Hallman JL, Hawkins PR;
XX	
DR	WPI: 1999-633321/54.
XX	
PT	P-PSDB: AAY41160.
XX	
PS	Human pyrophosphatase used in treatment and prevention of cancer -
XX	Disclosure; Fig 1A-D; 30pp; English.

XX The present sequence encodes a human inorganic pyrophosphatase
 CC designated HPYP. Nucleic acids encoding HPYP were first identified
 CC in Incyte clone 768320 from the lung tissue cDNA library LUNG0704.
 CC The recombinant pyrophosphatase, may be useful in promoting cell
 CC and tissue regeneration.

SO Sequence 1275 BP; 394 A; 235 C; 294 G; 351 T; 1 other;

Query Match 100.0%; Score 1275; DB 20; Length 1275;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGAGTTGGGAGCTCTCTGTCAGTCGGGCGCGGCTGGGGTGGTGGTCTGTG 60
 Db 1 caagagttgggagctctctgtcagtcgggcgcgcgctgggggtgggtggctgtg 60
 QY 61 GCAGCGGCGGCGGAGAGACTCCGCACTATGAGCGGCTTACAGCAGGAGCGCGC 120
 Db 61 gcagcgcgcgcgagagactccgcactatgagcggtttagcagcgagagcgcgcg 120
 QY 121 GCCCTTCTCCTGAGTACCGAGCTTCTTCAAAATGAGAAAGACATATATCTCC 180
 Db 121 gcccttctcctgagtagccgagcttctccaaatgagaaagacatatatctcc 180
 QY 181 ATTTCATGATATTCATTTATGAGATAGAGATGCTTTCACATGATGATGATACC 240
 Db 181 atttcgatatttcatttatgagatagagatgcttttcacatgatgatgataacc 240
 QY 241 ACCGTGCTTAATGCAAAATGAGATGCTTACAAAGACCTTTAAACCTTTAAACA 300
 Db 241 accgtgcttaatgcaaaatgagatgctttacaaagaccttttaacctttaaca 300
 QY 301 AGATGTGAAAAAGAAAACCTGCTATGTTGGCAATTTTCCGTATTAAGATATAT 360
 Db 301 agatgtgaaaaagaaaacctgctatgttggcaattttccgtatataagatatat 360
 QY 361 CTGGACTATGTCGATCCCTGCACTTGGGAAGGCCGACACATATATACTAC 420
 Db 361 ctggaactatgtcgatccctgcacttgggaagggcgacacatataactac 420
 QY 421 TGGCTGTGTGTGACAAATGACCAATGATGTGTGAAATTTGGAAGCAAGATATGC 480
 Db 421 tggctgtgtgtgacaaatgacccaatgattgtgtgaaatttggaagcaagatgtgc 480
 QY 481 AAGAGTGAATTAATTTGGCGTGAAGTTCTAGCATTGGCTATGATTAACGAGGGA 540
 Db 481 aagagtgaaatatttggcgtagaagttctagcattggctatgattgacgaaggga 540
 QY 541 AACCGATGGAAGTCAATTCCTTAATGATGATGATGATGATGATGATGATGATG 600
 Db 541 aaccgatggaaagtcaattccttaattgatgatgatgatgatgatgatgatgatg 600
 QY 601 TATCAATGATGTAAGAGCGCTGAACCTGCTACTTGAAGACTACTGTGACTGTTAG 660
 Db 601 tatcaatgatgtgaagcgctgaacctgctacttgaagactactgtgactgttag 660
 QY 661 AAGGTATAGAGTTCCTGATGAGAAACCAAGAAATGAGTTTGGCTTTAATGACA 720
 Db 661 aaggtatagagtctctgatgagaaaccaaagaaatgagtttggcttttaattgaca 720
 QY 721 AGATAGGAGTTCCTGATGAGAAACCAAGAAATGAGTTTGGCTTTAATGACA 780
 Db 721 agataggagtctctgatgagaaaccaaagaaatgagtttggcttttaattgaca 780
 QY 781 GACTAGAAAGCAATGAGAAAGCAATGAGTTCATGATGATGATGATGATGATGAT 840
 Db 781 gactagaaagcaatgagaaagcaatgagttcatgatacaacttgcctgagagccc 840
 QY 841 CTTCAAGTGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Db 841 ctccaagtgtgatactgatactgatactgatactgatactgatactgatactgata 900

QY 901 ATCTGCTGACAGTACCAACAGAGTGGATAGTGGTTCATCCAGAGAAACTAATG 960
 Db 901 atctgctgacagtagccaacagagtggtatgaagtggttccatccagaaaactaag 960
 QY 961 AGATTTCTGGAATFACAGCTGATATGTCATCGTGTTCATCTGATGATGATGATG 1020
 Db 961 agatttctggaaatfacaagctgatattgtcacaactgtgttcacatctgattagaag 1020
 QY 1021 TAAAGTAGTAGCTTTTCAAGCTTTTAAATTTGAGAACTCATTAAGTAAATTC 1080
 Db 1021 taaagtagtagcttttcaagctttttaaatttgagaaactcatagaaatgaaatc 1080
 QY 1081 TGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
 Db 1081 tgggtgtgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatg 1140
 QY 1141 ACTAAGTAACTTTTACACATGCTTAAATATCAACAGAGTGTTCATTTGGAAGTCA 1200
 Db 1141 actaagttaacttttaccatgctttaaattatcaaacagagtggttcatttgaagtcact 1200
 QY 1201 GTGAAATGATGTCGAAGGAGACACATTTGATGATGATGATGATGATGATGAT 1260
 Db 1201 gtgaaatgatgtcgaaaggagacacatttgatgattatgtatcattatgtaggaat 1260
 QY 1261 AAAATTTTTCGTCG 1275
 Db 1261 aaaatttttcgctg 1275

RESULT 3
 AAS08596
 ID AAS08596 standard; cDNA; 1277 BP.
 AC AAS08596;
 XX
 XX 24-OCT-2001 (first entry)
 DE
 XX DNA encoding human pyrophosphatase (HPYP).
 KW Pyrophosphatase; human; HPYP; antiinflammatory; cytosolic;
 KW immunosuppressive; antianaemic; antidiabetic; osteopathic;
 KW antineumatic; antiathrilitic; inflammatory disease; cancer; nerve;
 KW transpiration; pancreatic; bone marrow; prostate; anemia; asthma;
 KW systemic lupus erythematosus; myasthenia gravis; diabetes mellitus;
 KW osteoporosis; glomerulonephritis; rheumatoid arthritis; osteoarthritis;
 KW scleroderma; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 91..960
 FT /tag= a
 FT /product= "Human pyrophosphatase"
 US2001010911-A1.
 PD 02-AUG-2001.
 XX
 XX 08-OCT-1999; 99US-0415540.
 PF
 XX 31-OCT-1996; 96US-0741437.
 PR 13-AUG-1998; 98US-0134593.
 XX
 XX (HAWK) HAWKINS P R.
 PA (HILL) HILLMAN J L.
 PI Hawkins PR, Hillman JL;
 XX
 XX WPI: 2001-502139/55.
 DR P-SDB: AA005122.
 XX
 XX New human pyrophosphatase polypeptides and polynucleotides, useful for

PT diagnosing, preventing and treating inflammatory diseases and cancer,
 PT or in screening libraries of compounds for various drug screening
 techniques

PS Claim 5; Fig 1; 32pp; English.

XX
 CC The invention relates to novel human pyrophosphatase (HPP) polypeptides
 CC and polynucleotides. The HPP nucleic acids and amino acids are useful
 CC in the diagnosis, prevention and treatment of inflammatory diseases and
 CC cancer. HPP and the vector encoding HPP may be used to in the
 CC regeneration or transplantation and development of nerve, pancreatic,
 CC bone marrow, prostate tissue and could involve supplying various
 CC natural molecules, in drug regimes that allow cell division and
 CC differentiation to occur. The polynucleotides may be used to detect and
 CC quantify gene expression in biopsied tissues in which expression of
 CC HPP may be correlated with a disease, for screening libraries of
 CC compounds in various drug screening techniques, and in molecular
 CC biology techniques. Vectors expressing antisense and antagonists or
 CC inhibitors of the protein may be used to suppress the excessive
 CC proliferation of inflammatory cells which cause damage in immunological
 CC diseases, such as anaemia, asthma, systemic lupus erythematosus,
 CC myasthenia gravis, diabetes mellitus, osteoporosis, glomerulonephritis,
 CC rheumatoid and osteoarthritis, and scleroderma. The present sequence
 CC represents the coding sequence of human pyrophosphatase.

XX Sequence 1277 BP; 394 A; 235 C; 294 G; 351 T; 3 other;

Query Match 100.0%; Score 1275; DB 22; Length 1277;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGAGTGGGCGCTCTCTCTGTCAGTGGCGCCCGGCTGGTGGTGGTGGTGG 60
 Db 3 caagaggtlmggggct 62
 QY 61 GCAGGCGGCGGCGGCGAGAGCTCGGACATGAGCGGCTTACACAGAGAGAGCGCCG 120
 Db 63 gcagggggyggcgagagctcgacacatgagcggtcttcgacagggagcgcgccgc 122
 QY 121 GCCCTTCCCTCGAGTACCGAGTCTTCTCTCAAAAATGAGAAAGACAATATATCTCC 180
 Db 123 gcccttccctcgagagctctctctctctctctctctctctctctctctctctct 182
 QY 181 ATTTCATGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 240
 Db 183 atttcagatattccaattatgagataagatgtgtttacacatgagtagtgaagtag 242
 QY 241 ACAGTGTCTATGCAAAAATGAGATGCTACAAAGAGACCCCTTAAACCCATTAAACA 300
 Db 243 acagctgtctaatgcaaaaaatgagatgtgcaaaaaagacccctttaaaccctataaaca 302
 QY 301 AGATGTGAAAAAGAAAACCTTCCGTATGTCGCAATTTGTTCCCGTATTAAGATATAT 360
 Db 303 agatgtgaaaaaagaaaacttctgctatgttgcgaattgttcccgatataaagagatat 362
 QY 361 CTGGAACATGATGTCGATCCTCTGAGACTTGGAGAGACCCAGGACCATGATTAACATAC 420
 Db 363 ctggaactatgtgtccatccctcagacttggagagaccagggagcaatgataaacaacac 422
 QY 421 TGGCTGTTGGTGAACATGACCAATTTGATGTGTGTAATTTGGAACAGATATGTCG 480
 Db 423 tggctgttgggtgacaatgacccaattgtatgttgaatttgaagcaagatgtatgyc 482
 QY 481 AAGAGGTGAATTAATTTGGCGTGAAGATTCTAGCATATTTGGCTATGATGACGAGGGA 540
 Db 483 aagaggtgaataatattggtcgtaaaagtctagcataatgtatgtatgacgaaggaga 542
 QY 541 AACGAGCTGGAAAGTCAATTTGCATTATGTCGATGATTCCTATGACGACCATTTATAAGA 600
 Db 543 aacgagctggaaagatcattgcattatgtatgtatgtatgtatgtatgtatgtatgtat 602
 QY 601 TATCAATGATGTCAAGAGGCTGAACCTGGCTACTAGAACCTACTGTGAGCTGTTAG 660

Db 603 tatcaatgtatgataaacagctgaaacctgtctacttagaagcctatgtgactgttag 662
 QY 661 AAGGTAAAGGTTCCTGATGAGAAACACAGAAAATGAGTTGGCTTAAACAGAAATTTAA 720
 Db 663 aaggtataaggttccctgataagaaacagaaatgagttgttgcattaaagcaatttaa 722
 QY 721 AGATTAAGACCTTTGCCATTTGATTTTAAAGACCTCATGACCAATTTGAACCAATTA 780
 Db 723 agataagacattgtccatgtatattataaagcactcatgaccatttgaagcattag 782
 QY 781 GACTAGAAAGCAATGAGAAAGCATGCTTCATGATGATGATGATGATGATGATGATGAT 840
 Db 783 gactaagaaacgaaagaaagaaatcagttgcatgaaacaaacttcttgaagacc 842
 QY 841 CTTCAGGTGATTCCTGAGTCTGCGAGACCATTTGGATGCTTACACACACCTGTCGA 900
 Db 843 cttaagtgtagtccatgacgtgcgcagagcatatgtgtagtcttaccacaccctgtga 902
 QY 901 ATCTGCTTCACAGATACCAACAGACGTGATGATGATGATGATGATGATGATGATGAT 960
 Db 903 atctgcttcacagatcaccaacagacgtgataagtgatgttccatccagaaaaactaat 962
 QY 961 AGATTTCTGTGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 Db 963 agatttctgtgaataacaaagcgtatgtcatatcgtatcattcgtatgtatgaag 1022
 QY 1021 TAAAGTACTAGCTTTTCAAGCTTAAATTTGTAGACATCAATCAATCAATCAATCAAT 1080
 Db 1023 taaagtagtagcttcttcaagctttaaatttltgtagaactcatcctaaagtaattc 1082
 QY 1081 TCGTGTGATATCCATATCTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 Db 1083 tgcgtgtactaatccaatactacacagatgtatccatccatccatccatccatccatcc 1142
 QY 1141 ACTAAGATACCTTTAGACATGCTTAAATTTAAAGAGCTGATTTGGAAGTCACTT 1200
 Db 1143 actaagataactttagacacatgctttaaataacaaagcgttgcatttggaaagtaact 1202
 QY 1201 GTGAATAGATGCAAGGAGACACATATTTGATGATGATGATGATGATGATGATGATGAT 1260
 Db 1203 gtgaatagatgtgcaagggagcacatattgtatgtatgtatgtatgtatgtatgtat 1262
 QY 1261 AAAATTTATTTGCTG 1275
 Db 1263 aaattatttgcgt 1277

RESULT 4
 ID AAX00685 standard; DNA; 1351 BP.
 AAX00685;
 25-MAR-1999 (first entry)

Human secreted protein gene 75 clone H05E145.
 Human; secreted protein; fusion protein; gene therapy; protein therapy;
 diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;
 inflammatory; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.
 XX
 XX W09842738-A1.
 PN
 XX
 XX 01-OCT-1998.
 PD
 XX

QY 1271 TCCTG 1275
11111
Db 1280 Tgctg 1284

RESULT 5
AA000716
ID AAX00716 standard: DNA: 1350, BP.
XX
AC AAX00716;
XX
DT 25-MAR-1999 (first entry)
XX
DE Human secreted protein gene 75, clone H0SE145.
XX
KW Human: secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis: cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.
XX
PN WO9842738-A1.
XX
PD 01-OCT-1998.
XX
PF 19-MAR-1998; 98WO-US05311.
XX
PR 30-MAY-1997; 97US-0050937.
PR 21-MAR-1997; 97US-0041276.
PR 21-MAR-1997; 97US-0041277.
PR 21-MAR-1997; 97US-0041281.
PR 30-MAY-1997; 97US-0042344.
PR 30-MAY-1997; 97US-0048069.
PR 30-MAY-1997; 97US-0048094.
PR 30-MAY-1997; 97US-0048095.
PR 30-MAY-1997; 97US-0048096.
PR 30-MAY-1997; 97US-0048099.
PR 30-MAY-1997; 97US-0048131.
PR 30-MAY-1997; 97US-0048135.
PR 30-MAY-1997; 97US-0048154.
PR 30-MAY-1997; 97US-0048160.
PR 30-MAY-1997; 97US-0048186.
PR 30-MAY-1997; 97US-0048187.
PR 30-MAY-1997; 97US-0048188.
PR 30-MAY-1997; 97US-0048350.
PR 30-MAY-1997; 97US-0048351.
PR 30-MAY-1997; 97US-0048352.
PR 30-MAY-1997; 97US-0048355.
PR 05-AUG-1997; 97US-0054804.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
PI Rosen CA, Ruben SM, Shi Y, Young P;
XX
DR WPI: 1999-070066/06.
DR P-PSDB: AAM67912.
XX
PT New isolated human genes and the secreted polypeptides they encode
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1: Page 264-265; 385pp; English.
XX
CC This sequence represents a nucleic acid molecule which encodes a
CC secreted human protein. The gene number, and the clone it is derived

CC from, are detailed in the descriptor line. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. AAX00602) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 87 novel genes and their fragments (nucleic
CC acid sequences: AAX00611-X00724; amino acid sequences AAM67807-W68004)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 87
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX00611 for described uses).

XX
CC Sequence 1350 BP; 414 A; 250 C; 319 G; 363 T; 4 other;

Query Match 90.4%; Score 1153; DB 20; Length 1350;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CCTTCTCCCTGGAGTACCGAGCTTCTCCCAAAATGGAAGGACAAATATATCTCCAT 182
Db 131 ccttctccctggagtagcagagcttctcccaaaatggaaggaacataatactccat 190
QY 183 TTCATGATATTCATTTATGACAGTAAAGATGTTGTCACATGATGTAAGTACCAC 242
Db 191 ttcatgatattccatttatgacagtagagtagtgcacatgtagtgaagtagccac 250
QY 243 GCTGCTTAATGCAAAATGAGATGCTACAAAGACCTTTAAACCTTTAAACAG 302
Db 251 gctgcttaatgcaaaaatgagatgctacaaagaccctttaaacccttaaacag 310
QY 303 ATGTGAAAAAGAAACCTGCTATGCGAATTTGTCGATTAAGATATATCT 362
Db 311 atgtgaaaaagaaaccttgcctatgctgtaattgtccgataaagataatct 370
QY 363 GGAAGTATGCTGCTCCCTCAGACTTGGGAAGACCCAGGACCAATATATATCTG 422
Db 371 ggaactatgctgctccctcagacttgggaagaccagggacaaatgataaacatctg 430
QY 423 GCTGTTGGTGACAAATGACCAATGATGTTGTAATGGAATGGAAGCATGTGCA 482
Db 431 gctgttggtgacaaatgaccaaattgattggtgtaattggaaagcaagtagtgcac 490
QY 483 GAGTGAAATAATTTGGCGTGAAGTTCTAGGCAATTTGGCTATGATGACGAAGG 542
Db 491 gagtgaaataatttggcggtgaagttctagacataatgtgataatgacgaagggaaa 550
QY 543 CCGACTGGAAGTCAATGCTTATATGATGATGATGATGATGATGATGATGATG 602
Db 551 ccgactggaaagtcattgcatlaatgtagatccgtagcagcaacatataatgata 610
QY 603 TCAATGATGTCAAACGGCTGAACCTGCTACTTGAAGCTACTGTGACTGTGTTA 662
Db 611 tcaatgatgtcAAACGGCTGAACCTGCTACTTGAAGCTACTGTGACTGTGTTA 670
QY 663 GGTATAAGTTCCTGATGGAAGAAACAGAAATAGCTTTCCTTTAATGCAATTTA 722
Db 671 ggtataagttcctgatggaaagaaacagaaatagcttcttcaatgcaatTTA 730
QY 723 ATAAAGCACTTCCCATGATTAATTAAGCAATGATGATGATGATGATGATG 782
Db 731 ataaagcacttcccatgattatataaagcaatgacatgacatggaagaagataag 790
QY 783 CTAAAGAAAGCAATGAAAGCAATGATGATGATGATGATGATGATGATGATG 842
Db 791 ctaaagaaagcaatgaaagcaatgacatgacatgacatgacatgacatgacatg 850
QY 843 TCAAGTGTATCTGATGCTGCTGCAAGGACCATGTTGATGATGATGATGATG 902
Db 851 tcaagtgtatctgatgctgctgcaagacatgacatgacatgacatgacatgacat 910

QY 903 CTGCTGACAGTACCAACAGACGTGATAAGTGTTCATCCACCAGAAAACTAATGAG 962
|||||
Db 911 ctgctgacagctacacacagagctgataagtggtccatccaccagaaaactaatgag 970
QY 963 ATTTCTGTGAATACAAAGCTGATATTCCTACATCGTGTTCATCTGATGTAAGAATA 1022
|||||
Db 971 attctctggaatacaagctgatatgtcatatcgtgttcacgtatgtagaagta 1030
QY 1023 AAGTACTAGCTTTTCAAGCTTAAATTTGTAGACATCTAAGCAAGTAAATCTG 1082
|||||
Db 1031 aaagtacagcttccaaagctttaaattgtagaactcaactaaagtaaatcttg 1090
QY 1083 CTGTGATTAATCCAAATTAATCTAGATATTCATCTAAGCAAGTAAATCTGCAAC 1142
|||||
Db 1091 ctggaactaatccaatatactacagaaatgtaatacctaaagcaatttcatactcaac 1150
QY 1143 TAAGTAACTTTTGGACATGCTTAATATCAAAAGCAAGTGTCTTTTGAAGTACTTGT 1202
|||||
Db 1151 taagataacttctgacacatgcttaaaatacaagcagctgtcatctggaagtaactgt 1210
QY 1203 GAATAGATGTGCAAGGAGACATATTTGATGTATGTATGATCAATATGTTAGAAATTA 1262
|||||
Db 1211 gaataagctgcaagggagacacatgtgataatgtatgtaccataatgttagaataa 1270
QY 1263 AATTATTTTCTGCTG 1275
|||||
Db 1271 aattatttctg 1283

RESULT 6

AAZ33425 standard; cDNA: 1200 BP.

ID AAZ33425;

AC AAZ33425;

XX 08-DEC-1999. (first entry)

DE Human prostate cancer-associated EST 3.

KM Expressed sequence tag: EST; prostate tumor; antitumor; treatment;

XX gene therapy; tissue specificity human; ss.

XX Homo sapiens.

OS DE19811193-A1.

PD 16-SEP-1999.

PF 10-MAR-1998; 98DE-1011193.

PR 10-MAR-1998; 98DE-1011193.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;

PI WPI: 1999-519628/44.

DR P-PSDB: AAY48218.

XX New nucleic acid expressed at high level in prostatic tumor tissue and

PT encoded polypeptides, useful for treating cancer and screening for

XX therapeutic agents

XX Claim 1a; 71: 166pp; German.

XX This invention describes novel nucleic acid sequences (A) that are

CC expressed at high level in prostatic tumor tissue and encode gene

CC products or their fragments. The products of the invention have

CC antitumor activity. Polypeptides (I) encoded by (A) are used: (i) for

CC identifying agents for treatment of prostatic cancer and (ii) for

CC therapy of prostate cancer, optionally where expressed by gene therapy

CC methods. (A) is also used to isolate full-length genes (for gene therapy)

CC and for recombinant production of (I), which can be used to raise

CC specific antibodies. (A) are identified by assembly of ESTs (expressed
CC sequence tags) before they are analyzed for expression pattern (tissue
CC specificity). This approach eliminates many of the false results, as
CC regards tissue specificity, associated with known methods that use
CC single (usually short) ESTs. AAZ33423-433476 represent expressed
CC sequence tags described in the method of the invention.

XX Sequence 1200 BP; 368 A; 227 C; 278 G; 327 T; 0 other;

Query Match 82.7%; Score 1055; DB 20; Length 1200;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1055; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CCTTCTCCCTGAGTACGAGCTTCTCCCAAAATGAGAAAGCAATATATCTCAT 182
|||||
Db 111 ccttctccctgagatcagagcttcccaaaaagaaagaaatataatctccat 170
QY 183 TTCAATGATATTCATATTTATGCAATAGAGATGTTTACATGCTAGTGAATACGAC 242
|||||
Db 171 ttcataatattccaatattatgcagataagatggtttccatcatgtagtgaagtaacac 230
QY 243 GCTGTCTAATGCAAAATGAGATGCTGATCAAGAGACCTTTAAACCTTTAAACAG 302
|||||
Db 231 gctgtctcaatgcaaaaatgagatgcttcaaaaagacctttaaaccctataaacaag 290
QY 303 ATGTGAAAAAGAAACCTTGCTATGTTGCAATTTGTCCTCCGATTAAGATATATCT 362
|||||
Db 291 atgtgaaaaaagaaaccttgcataatgctgaatgttcccgatataaagataatct 350
QY 363 GGAATATGTCGATCCCTCAGACTTGGAGAACCCAGGACCAATGATTAACATAGTACT 422
|||||
Db 351 ggaactatgctcaccctcagacttgggaagaccagggacaatataaactactg 410
QY 423 GCTGTGTGTCGACATATCCCATTTGATGCTGTGGAATTTGGAAGCAAGTATGCA 482
|||||
Db 411 gctgtgtgtgacaaatgacccaatgtagtgtgtaaatggaagcaaggtatgtgcaa 470
QY 483 GAGTGAATATATTTGGCTGAAAGTGTAGCATTTGCTATGATTAAGCAAGGGGAA 542
|||||
Db 471 gagtgaaatattggtgtgaaagttccaaagcatatggtcatgtgacgaagggaaa 530
QY 543 CCGACTGGAAGATCATTTGCCATTAATGTGATGATGATGATGATGATGATGATGATG 602
|||||
Db 531 ccgactggaagatcatattgcatatgtagatgatacccgatgagcgaatataatgata 590
QY 603 TCAATGATGCAAAAGGCTGAACCTGGCTACTTGAAGAGTACTGTGAGCTGTTAGAA 662
|||||
Db 591 tcaatgattcaaaagctgaaacctgtacttgaagactgtgacgtgtttagaa 650
QY 663 GGTATTAAGCTTCCTGATGGAAGAACGAAATGAGTTGGCTTAATCCAGATTAAAG 722
|||||
Db 651 ggtataagcttccgtgaaagaaacgaagaatgagtttggtttaaagcaaatltaag 710
QY 723 ATAAGACTTTGCCATGATATTAATTAAGACATCATGACCATTTGGAAGCATTAAGTA 782
|||||
Db 711 ataagactttgcatattatatttaaaagcaccatgacatattggaagcattagaga 770
QY 783 CTAAGAAAGCAATGGAAGAGATCACTGATGATTAACAATTTGCTGAGAGCCCT 842
|||||
Db 771 ctaagaaacgaatggaagaaagatcagttgcatbaatacaacttgtctgagagccct 830
QY 843 TCAAGTGTGATCTGATGCTGATGCTGCAAGGACATTTGATGCTTACCACCCCTGTGAT 902
|||||
Db 831 tcaagtgatctcgtgagctgcccagagccatgtgagatgcttaccaccacctgtgat 890
QY 903 CTGCTGACAGTACCAACAGACGTGATAAGTGTTCATCCACCAGAAAACTAATGAG 962
|||||
Db 891 ctgctgacagctacacacagagctgataagtggttccatccaccagaaaactaatgag 950
QY 963 ATTTCTGTGAATACAAAGCTGATATTCCTACATCGTGTTCATCTGATGTAAGAATA 1022
|||||
Db 951 attctctggaatacaagctgatatgtcatatcgtgttcacgtatgtagaagta 1010

[illegible]

OS Homo sapiens.
XX EPI033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-0200610.
XX 26-FEB-1999; 99US-0122487.
XX (GEST) GENSET.
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI: 2000-500381/45.
DR
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 14703; 71bp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
CC
SQ Sequence 660 BP; 218 A; 112 C; 133 G; 197 T; 0 other:

Query Match 51.7%; Score 659; DB 21; Length 660;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 617 CGGGTGAACCTGGCTACTGATGAGCTGCTGAGCTGTTTAAAGGTATAGTCTCT 676
Db 1 cggctgaacctgctacttaagaagctacgctgagctggtttagaaggctaaagcttcc 60
OY 677 GATGGAACCCAGAAATGATGCTTGCCTTATGACAAATTTAAAGATAGGACTTGC 736
Db 61 gatggaacccagaaatgagcttgcctttaaagcagatttaagaagacttcc 120
OY 737 ATTGATATTATTAAGACATGACCATGGAAGCATTTAGTGACTAGAAAAAGCAT 796
Db 121 atgtatattataaaagacatcagcattggaagcattagtgactaagaagaacgat 180
OY 797 GGAAGAGAAATCATGATGATACATCACTTGTCTGAGACCCCTTAAAGTGTGATCT 856
Db 181 ggaagagaaatcagtgatgaatacaacttgcctgagagcccttcaagtgtatcct 240
OY 857 GATGCTGCACAGCATTTGATGCTTACCAACCCCTGGAATCTGCTGCACATA 916
Db 241 gatgctgcacagcatTTGATGCTTACCAACCCCTGGAATCTGCTGCACATA 300
OY 917 CCAACAGAGCTGATTAAGTGGTTCATACCAAGAAAACTATGATTTCTGTGATA 976
Db 301 ccaacagagctgataaagtggttccatcaccaagaataatgagatttccctggaata 360
OY 977 CAAGCTGATTTGCTACCTGCTTTCATCTGATGATTTAGAAAGTAAGTAGTACTTT 1036
Db 361 caagctgatcttctacactgctgttcacatcgtatgtatttagaagaataagtagtctt 420
OY 1037 TCAAGCTTTAAATTTGTAGAAACATCATCTAACTAAAGTAATTCCTGCTGACTAATCA 1096

Db 421 tcaagctttaaattttagaactcatcctaactaagaataatcgcgtgactaatcca 480
OY 1097 ATATCTCAGAAATGTTATCCATCTTAACACATTTTTCATATCTCACTAAGTAAGTCTTA 1156
Db 481 atatactcagaatglttaccatcctaagaacttccatccactcaagaataacttca 540
OY 1157 GCACATGCTTAAATATCAAGACAGTTGTCATTTGGAAGTCACTTGTGAATAGATGCGAA 1216
Db 541 gcacatgctttaaataatcctaagaagcagttgcatcttggaaatcacttggatgattgcaa 600
OY 1217 GGGGAGCACATTTGGATGATGATGATACCATATGTTAGAAATAAATATTTGCTG 1275
Db 601 ggggagcacatatgtgattgatatgttaccatatgttgaataaataatttgcg 659

RESULT 9
AAA/8135/c
ID AAA/8135 standard; cDNA: 578 BP.
XX
AC AAA/8135;
XX
DT 14-NOV-2000 (first entry)
XX
DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:422.
XX
KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
KW Immunotherapy; diagnosis; progression; ss.
XX
OS Homo sapiens.
XX
PN WO200037643-A2.
XX
XX 29-JUN-2000.
PD
XX 23-DEC-1999; 99WO-US30909.
PF
XX 23-DEC-1998; 98US-0221298.
PR 02-JUN-1999; 99US-0347496.
PR 22-SEP-1999; 99US-0401064.
PR 19-NOV-1999; 99US-0444242.
PR 02-DEC-1999; 99US-0454150.
PA
XX (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk J;
PI Wang T, Yuguin J;
XX
DR WPI: 2000-442671/38.
XX
PT New colon tumor polypeptides used to inhibit the development of cancer,
PT especially colon cancer, and for diagnosing and monitoring the
PT progression of the cancer -
XX
PS Claim 1; Page 212; 229pp; English.
XX
CC Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
CC portions of proteins which are associated with human colon tumours.
CC The invention also specifically discloses 8 human colon tumour proteins
CC (AAB11897-811904). The nucleic acids, the polypeptides they encode, and
CC antigen presenting cells (APCs, preferably dendritic cells) expressing
CC such polypeptides may be used in vaccines that target tumour cells,
CC especially colon tumour cells, thereby inhibiting the development of
CC cancer. T-cells specific for the polypeptide expressed by the APC are
CC used to remove tumour cells from biological samples, especially blood or
CC fractions thereof. The sample or the isolated T-cells specific for the
CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
CC CD8+ T-cells from a patient may be incubated with a polypeptide or
CC nucleic acid of the invention, or an APC expressing such a polypeptide,
CC to cause the proliferation of specific T-cells. The T-cells can be
CC cloned and then administered back to the patient to inhibit cancer
CC development. Nucleic acids encoding the polypeptides and antibodies
CC against the polypeptides may be used to determine the expression level

AAT21836
ID AAT21836 standard; cDNA to mRNA; 416 BP.
AC AAT21836;
XX
DT 01-AUG-1996 (first entry)
XX
DE Human gene signature HMG503377.
XX
KW Gene signature: messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN M09514772-A1.
XX
PD 01-JUN-1995.
XX
PE 11-NOV-1994; 94MO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATSU) MATSUBARA K.
PA (OKUBA) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI: 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 977; 2245bp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-726837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 416 BP; 128 A; 77 C; 77 G; 128 T; 6 other;
XX
Query Match 24.9%; Score 317; DB 16; Length 416;
Best Local Similarity 99.7%; Pred. No. 8.1e-149;
Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
OY 851 GATCCTGATGCTGCAGACCATTTGTGATGCTTTACACACACCTGTGAATCTGCTGC 910
DB 1 gatccgtatgctgcagagccatgtgtgattgattacacacccctgtgactcgtcgc 60
OY 911 ACAGTACCAACAGACGTGATGATGCTTCCATCCACAGCAAAAATAATGATTTCT 970
DB 61 acagtaaccaacagacgtgataaagtggtccatccacgaataaactaatgatttct 120
OY 971 GGAATACAAGCTGATATGCTACATCGTTTCATCTGATTTAGAGTAAGTAAGTAGT 1030
DB 121 ggaatacaagctgatattgctacacatcggttcatctgattgattagaagaatagtg 180
OY 1031 AGCTTTCAAGCTTTAAATTTGTAGAACCTCATCTAACTAAAGTAATTTCTGCTGACT 1090

DB 181 agctttcaaaagctttaaatttgtagaactcatcctaagaagtaattcgtgact 240
OY 1091 AATCAATATACTAGATGATATCCATCAAGCATTTTTCATATCCACTAAGATTA 1150
DB 241 aatccaataactacgaagatgattccatcctaagaatttccatcctaagaataa 300
OY 1151 CTTTTCAGACATGCTTAAATATCAAGACAGTGTGATTTTGAAGTCACTTGTGAATAGAT 1210
DB 301 ctttagacacatgcttaataatcacaagcagtgcttattggaagcacttggaataag 360
OY 1211 GTGCAAG 1218
DB 361 gtgcaag 368
XX
RESULT 12
AAZ77472
ID AAZ77472 standard; cDNA; 357 BP.
XX
AC AAZ77472;
XX
DT 10-APR-2000 (first entry)
XX
DE Human ovarian tumor cDNA library derived EST fragment 23.
XX
KW Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW gene therapy; treatment; ss.
XX
OS Homo sapiens.
XX
PN DE19817557-A1.
XX
PD 21-OCT-1999.
XX
PE 09-APR-1998; 98DE-1017557.
XX
PR 09-APR-1998; 98DE-1017557;
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitz A, Pilarsky C, Dahl E;
XX
DR WPI: 1999-591920/51.
XX
P-PSDB: AAY76554, AAY76555.
XX
PT New nucleic acid sequences expressed in ovarian, and some other, cancer
PT tissues, and derived polypeptides, for treatment of ovarian cancer and
PT identification of therapeutic agents
XX
PS Claim 3; Page 157; 310pp; German.
XX
CC This invention describes novel nucleic acid (cDNA) sequences (A) which
CC have anticancer activity and are highly expressed in ovarian tumor
CC tissue (and some also in testis and breast cancer tissue). The products
CC of the invention can be used for gene therapy. (A) are used (i) for
CC recombinant expression of polypeptides (B) and (ii) to isolate complete
CC genes. (B) are used (i) to identify agents suitable for treatment of
CC ovarian cancer; (ii) directly for treating this form of cancer
CC (including expression from gene therapy vectors) and (iii) for generation
CC of specific antibodies. (A) are identified by assembling ESTs (expressed
CC sequence tags) from a particular tissue type before comparison of
CC expression patterns. This allows a significantly longer fragment of the
CC gene to be revealed, so should reduce the number of failures associated
CC with the fact that ESTs from different libraries may represent different
CC parts of the same unknown gene, distorting the estimated frequency of
CC occurrence in a particular tissue. AAZ77450-277572 represent the human
CC ovarian tumor cDNA library derived EST fragments described in the method
CC of the invention and encode the protein fragments represented in
CC AAY6505-Y76638.
XX
SQ Sequence 357 BP; 74 A; 89 C; 96 G; 98 T; 0 other;

CC cytosolic; antibacterial; antifungal; antiviral; antidiabetic;
CC antistimatic; vulnereary; antitumor; osteopathic; neuroprotective;
CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; antidepressant. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumors, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 102 BP; 34 A; 17 C; 27 G; 24 T; 0 other;

Query Match 5.38; Score 68; DB 21; Length 102;
Best Local Similarity 100.0%; Pred. No. 4.5e-24;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 407 AATGATTAACATCTGCTGTCGTGACATGACCAATTCATGTCGTAATTGGA 466
Db 27 aatgataaaacatactgctgtgtgtgacaaagaccatgtgtgtgaattgga 86
OY 467 AGCAAGCT 474
Db 87 agcaagct 94

RESULT 15
AA122766/c
ID AA122766 standard; DNA; 186 BP.
XX
AC AA122766;
DT 12-OCT-2001 (first entry)
XX
DE Probe #12699 for gene expression analysis in human cervical cell sample.
XX
KM Probe: human; microarray; gene expression; cervical epithelial cell;
XX
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN MO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 12699; 487bp; English.

XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human Hela cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 186 BP; 47 A; 44 C; 33 G; 62 T; 0 other;

Query Match 5.28; Score 66; DB 22; Length 186;
Best Local Similarity 100.0%; Pred. No. 4.5e-23;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 326 TATGTCGAATTGTTCCGATATAAGATATATGAGACTATGTCGTCATCCCTCAG 385
Db 147 TATGTCGAATTGTTCCGATATAAGATATATGAGACTATGTCGTCATCCCTCAG 88
OY 386 ACTTGG 391
Db 87 ACTTGG 82

RESULT 16
AA148068/c
ID AA148068 standard; DNA; 186 BP.
XX
AC AA148068;
DT 17-OCT-2001 (first entry)
XX
DE Probe #16754 used to measure gene expression in human placenta sample.
XX
KM Probe: microarray; human; placenta; antenatal diagnosis;
XX
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN MO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 16754; 654bp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX Sequence 186 BP; 47 A; 44 C; 33 G; 62 T; 0 other;
SQ

Query Match
Best Local Similarity 100.0%; Score 66; DB 22; Length 186;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 147 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAACATATGTCGCAATCCCTCAG 88

QY 326 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAACATATGTCGCAATCCCTCAG 385

DB 147 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAACATATGTCGCAATCCCTCAG 88

QY 386 ACTTGG 391

DB 87 ACTTGG 82

RESULT 17
AA108439/C
ID AA108439 standard; DNA; 186 BP.

XX AA108439;
AC
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #8430 used to measure gene expression in human breast sample.
XX
KW Probe: human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN W0200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast
XX
PS Claim 25; SEQ ID No 8430; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease, and
XX non-carcinoma tumours.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 186 BP; 47 A; 44 C; 33 G; 62 T; 0 other;
SQ

XX Query Match
Best Local Similarity 100.0%; Score 66; DB 22; Length 186;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 147 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAACATATGTCGCAATCCCTCAG 88

QY 326 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAACATATGTCGCAATCCCTCAG 385

DB 147 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAACATATGTCGCAATCCCTCAG 88

QY 386 ACTTGG 391

DB 87 ACTTGG 82

RESULT 18
AA113542/C
ID AA113542 standard; DNA; 411 BP.

XX AA113542;
AC
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #3475 for gene expression analysis in human cervical cell sample.
XX
KW Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN W0200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-489901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells
XX
PS Claim 25; SEQ ID No 3475; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
XX (SENP). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;
SQ

XX Query Match
Best Local Similarity 100.0%; Score 66; DB 22; Length 411;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAACATATGTCGCAATCCCTCAG 385

```

Db      358 TATGTCGGAATTTGTCCTGATTAAGATATATCTGGAACATATGTCCTCCTCAG 299
QY      386 ACTTG 391
        |||||
Db      298 ACTTG 293

RESULT 19
AAI34904/C
ID      AAI34904 standard; DNA: 411 BP.
AC      AAI34904;
XX      17-OCT-2001 (first entry)
XX      DE
XX      Probe #3590 used to measure gene expression in human placenta sample.
XX      KW
XX      Probe: microarray; human; placenta; antenatal diagnosis;
XX      genetic disorder; ss.
XX      OS
XX      Homo sapiens.
XX      PN
XX      WO200157272-A2.
XX      PD
XX      09-AUG-2001.
XX      PF
XX      30-JAN-2001; 2001WO-US00663.
XX      PR
XX      04-FEB-2000; 2000US-0180312.
XX      PR
XX      26-MAY-2000; 2000US-0207456.
XX      PR
XX      30-JUN-2000; 2000US-0608408.
XX      PR
XX      03-AUG-2000; 2000US-0632366.
XX      PR
XX      21-SEP-2000; 2000US-0234687.
XX      PR
XX      27-SEP-2000; 2000US-0236359.
XX      PR
XX      04-OCT-2000; 2000GB-0024263.
XX      PA
XX      (MOLE-) MOLECULAR DYNAMICS INC.
XX      PI
XX      Penn SG, Hanzel DK, Chen W, Rank DR;
XX      DR
XX      WPI; 2001-48897/53.
XX      PT
XX      Human genome-derived single exon nucleic acid probes useful for
XX      analyzing gene expression in human placenta.
XX      PS
XX      Claim 25; SEQ ID No 3590; 654pp; English.
XX      CC
XX      The present invention relates to single exon nucleic acid probes (SENP).
XX      CC
XX      The present sequence is one such probe. The probes are useful for
XX      CC
XX      producing a microarray for predicting, measuring and displaying gene
XX      CC
XX      expression in samples derived from human placenta. The probes are useful
XX      CC
XX      for antenatal diagnosis of human genetic disorders.
XX      SO
XX      Sequence 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 66; DB 22; Length 411;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      326 TATGTCGGAATTTGTCCTGATTAAGATATATCTGGAACATATGTCCTCCTCAG 385
        |||||
Db      358 TATGTCGGAATTTGTCCTGATTAAGATATATCTGGAACATATGTCCTCCTCAG 299
QY      386 ACTTG 391
        |||||
Db      298 ACTTG 293

RESULT 20
AAI03431/C
ID      AAI03431 standard; DNA: 411 BP.

```

```

XX      AC
XX      AAI03431;
XX      DE
XX      09-OCT-2001 (first entry)
XX      DE
XX      Probe #3422 used to measure gene expression in human breast sample.
XX      KW
XX      Probe; human; breast disease; breast cancer; development disorder; ss;
XX      inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX      OS
XX      Homo sapiens.
XX      PN
XX      WO200157270-A2.
XX      PD
XX      09-AUG-2001.
XX      PF
XX      29-JAN-2001; 2001WO-US00661.
XX      PR
XX      04-FEB-2000; 2000US-0180312.
XX      PR
XX      26-MAY-2000; 2000US-0207456.
XX      PR
XX      30-JUN-2000; 2000US-0608408.
XX      PR
XX      03-AUG-2000; 2000US-0632366.
XX      PR
XX      21-SEP-2000; 2000US-0234687.
XX      PR
XX      27-SEP-2000; 2000US-0236359.
XX      PR
XX      04-OCT-2000; 2000GB-0024263.
XX      PA
XX      (MOLE-) MOLECULAR DYNAMICS INC.
XX      PI
XX      Penn SG, Hanzel DK, Chen W, Rank DR;
XX      DR
XX      WPI; 2001-476286/51.
XX      PT
XX      Novel single exon nucleic acid probe used to measuring gene expression
XX      in a human breast.
XX      PS
XX      Claim 25; SEQ ID No 3422; 322pp; English.
XX      CC
XX      The present invention relates to novel single exon nucleic acid probes.
XX      CC
XX      The present sequence is one such probe. The probes are useful for
XX      CC
XX      measuring human gene expression in a human breast sample, where the probe
XX      CC
XX      hybridises at high stringency to a nucleic acid expressed in the human
XX      CC
XX      breast. The probes are useful for predicting, diagnosing, grading,
XX      CC
XX      staging, monitoring and prognosing diseases of the human breast,
XX      CC
XX      particularly those diseases with polygenic aetiology. The diseases
XX      CC
XX      include: breast cancer, disorders of development, inflammatory diseases
XX      CC
XX      of the breast, fibrocystic changes, proliferative breast disease and
XX      CC
XX      non-carcinoma tumours.
XX      CC
XX      Note: The sequence data for this patent did not form part of the printed
XX      CC
XX      specification, but was obtained in electronic format directly from WIPO
XX      CC
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX      SO
XX      Sequence 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 66; DB 22; Length 411;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      326 TATGTCGGAATTTGTCCTGATTAAGATATATCTGGAACATATGTCCTCCTCAG 385
        |||||
Db      358 TATGTCGGAATTTGTCCTGATTAAGATATATCTGGAACATATGTCCTCCTCAG 299
QY      386 ACTTG 391
        |||||
Db      298 ACTTG 293

RESULT 21
AAH35811
ID      AAH35811 standard; CDNA: 342 BP.
AC      AAH35811;
XX

```


RESULT 23
AAZ52867 standard; cDNA: 1202 BP.
XX
XX AAZ52867;
AC
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor cDNA library derived EST fragment #10.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytosolic;
XX treatment; ds.
XX
XX Homo sapiens.
OS
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
PA
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
PI
XX WPI: 1999-621386/54.
DR P-PSDB; AAY73841, AAY73842, AAY73843.
XX
XX New human nucleic acid sequences from pancreatic tumors, and related
PT proteins
XX
XX Claim 2; Page 189-190; 502pp; German.
PS
XX This invention describes novel polypeptides and their encoding nucleic
CC acids derived from human pancreatic tumor tissue which have cytosolic
CC activity. The sequences are also useful in producing pharmaceutical
CC compositions for treatment of pancreatic tumors. AAZ52858-253014
CC represent expressed sequence tag (EST) fragments derived from a human
CC pancreatic tumor cDNA library and which encode the proteins represented
CC in AAY73814-74252.
CC
XX
SQ Sequence 1202 BP; 381 A; 227 C; 267 G; 327 T; 0 other;

Query Match 1.5%; Score 19; DB 20; Length 1202;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 TCCCTCAGACTTGGGAGA 396
Db 454 tccctcagacttgggaaga 472

RESULT 24
AAC98069
ID AAC98069 standard; cDNA: 1215 BP.
XX
AC AAC98069;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen nucleotide sequence SEQ ID NO:79.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytosolic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.
XX

OS Homo sapiens.
XX
XX WO200055351-A1.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 08-MAR-2000; 2000WO-US05883.
PF
XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
PI
XX WPI: 2000-587534/55.
DR P-PSDB; AAB53312.
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer
XX
XX Claim 1; Page 524; 2104pp; English.
PS
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytosolic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772; and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
XX
SQ Sequence 1215 BP; 381 A; 232 C; 273 G; 329 T; 0 other;

Query Match 1.5%; Score 19; DB 21; Length 1215;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 TCCCTCAGACTTGGGAGA 396
Db 454 tccctcagacttgggaaga 472

RESULT 25
AAI59678
ID AAI59678 standard; cDNA: 1536 BP.
XX
AC AAI59678;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 3667.
XX
KW Human; nootropic; immunosuppressant; cytosolic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX Homo sapiens.
OS
XX WO200153312-A1.
PN

PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
XX
PR 25-APR-2000; 2000US-0552317.
XX
PR 09-JUL-2000; 2000US-0598042.
XX
PR 19-JUL-2000; 2000US-0620312.
XX
PR 03-AUG-2000; 2000US-0653450.
XX
PR 14-SEP-2000; 2000US-0662191.
XX
PR 19-OCT-2000; 2000US-0693036.
XX
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
XX
DR P-PSDB: AAM40522.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 3667; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytoskeletal activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
SQ Sequence 1536 BP; 443 A; 365 C; 399 G; 329 T; 0 other;

Query Match 1.5%; Score 19; DB 22; Length 1536;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGGGCGGCGCAG 76.
|||
Db 346 gtggcagcgggcgggcgag 364

RESULT 26
ID AAI59679 standard; cDNA: 1536 BP.
XX
AC AAI59679;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 3668.
XX
XX Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX

OS Homo sapiens.
XX
XX W0200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX
XX 21-JAN-2000; 2000US-0488725.
XX
XX 25-APR-2000; 2000US-0552317.
XX
XX 09-JUL-2000; 2000US-0598042.
XX
XX 19-JUL-2000; 2000US-0620312.
XX
XX 03-AUG-2000; 2000US-0653450.
XX
XX 14-SEP-2000; 2000US-0662191.
XX
XX 19-OCT-2000; 2000US-0693036.
XX
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX
XX P-PSDB: AAM40523.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX
XX Claim 1; SEQ ID NO 3668; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytoskeletal activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 1536 BP; 443 A; 365 C; 399 G; 329 T; 0 other;

Query Match 1.5%; Score 19; DB 22; Length 1536;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGGGCGGCGCAG 76
|||
Db 346 gtggcagcgggcgggcgag 364

RESULT 27
ID AAV81267 standard; cDNA: 2264 BP.
XX
AC AAV81267;
XX
DT 11-MAR-1999 (first entry)
XX
XX Bcl-2 interaction protein encoding cDNA.
XX Bcl-2 interaction protein; Bcl-2; molecular mechanism; apoptosis;
XX diagnosis; prevention; treatment; ss.
XX

OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 213..1742
FT /*tag= a
FT /product= "Bcl-2 interaction protein"
XX
XX JP10327872-A.
XX
XX 15-DEC-1998.
XX
XX 30-MAY-1997; 97JP-0157841.
XX
XX 30-MAY-1997; 97JP-0157841.
XX
XX (SAKA) OTSUKA PHARM CO LTD.
XX
XX WPI: 1999-099033/09.
XX
XX P-PSDB: AAW95196.
XX
XX New Bcl-2 interaction protein gene (Bis) - useful for elucidation of
PT the molecular mechanism of apoptosis, and in diagnosis, prevention
PT and treatment of diseases
XX
XX Claim 2; Page 12-15; 19pp; Japanese.
XX
XX This cDNA encodes a Bcl-2 interaction protein which has combining
CC activity to Bcl-1. The Bcl-2 interaction protein and its gene (Bis) are
CC useful for the elucidation of the molecular mechanism of apoptosis and
CC various diseases related to it, and their diagnosis, prevention and
CC treatment.
XX
XX Sequence 2264 BP; 565 A; 687 C; 574 G; 437 T; 1 other;

Query Match 1.5%; Score 19; DB 20; Length 2264;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 58 GTGGCAGCGCGCGCGCAG 76
|||||
DB 474 gtggcagcgcgcgcgcgag 492

RESULT 28
AAZ51797
ID AAZ51797 standard; cDNA: 2528 BP.
XX
XX AAZ51797;
XX
XX 04-JUL-2000 (first entry)
XX
XX Full length cDNA sequence for human BAG-3 protein.
XX
XX Human BAG-3; Bcl-2 associated athanogene-3; apoptosis; cell migration;
KW tumour cell metastasis; Hsc70/Hsp70-regulating protein; metastasis;
KW tumour cell proliferation; steroid hormone receptor function; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 1..2034
FT /*tag= a
FT /product= "Human BAG-3 protein"
XX
XX WO200014106-A1.
XX
XX 16-MAR-2000.
XX
XX 09-SEP-1999; 99WO-US21053.
XX
XX 09-SEP-1998; 98US-0150489.
XX

PA (BURN-) BURNHAM INST.
XX
XX Reed JC, Takayama S;
XX
XX WPI: 2000-256937/22.
XX
XX P-PSDB: AAY70508.
XX
XX BAG-1 related proteins from humans, Caenorhabditis elegans and
PT Schizosaccharomyces pombe useful for modulating tumor cell
PT proliferation, cell migration and metastasis and steroid hormone
PT receptor function
XX
XX Claim 7; Fig 3; 132pp; English.
XX
XX The present cDNA sequence encodes human BAG-3 (Bcl-2 associated
CC athanogene-3) protein. BAG is a Hsc70/Hsp70-regulating protein
CC (Hsc70/Hsp70 is a molecular chaperone that participates in controlling
CC protein bioactivity, degradation, complex assembly/disassembly and
CC translocation across membranes). It competes with Hip for binding to the
CC Hsc70/Hsp70 ATPase binding domain and promotes substrate release. Gene
CC transfection studies indicate that BAG proteins influence a wide variety
CC of cellular phenotypes through their interactions with Hsc70/Hsp70,
CC including increasing resistance to apoptosis, promoting cell
CC proliferation, enhancing tumour cell migration and metastasis and
CC altering transcriptional activity of steroid hormones. The ability of
CC BAG-2 and BAG-3 proteins to interact in cells with Hsc70 was tested by
CC expressing these proteins with N-terminal Flag group tags in 293T human
CC epithelial cells using immunoprecipitation assays. BAG also stimulates
CC Hsc70-mediated adenosine triphosphate (ATP) hydrolysis by accelerating
CC ADP/ATP exchange.
XX
XX Sequence 2528 BP; 586 A; 805 C; 656 G; 478 T; 3 other;

Query Match 1.5%; Score 19; DB 21; Length 2528;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGCGCGCGCAG 76
|||||
DB 766 gtggcagcgcgcgcgcgag 784

RESULT 29
AAZ51804
ID AAZ51804 standard; cDNA: 2534 BP.
XX
XX AAZ51804;
XX
XX 04-JUL-2000 (first entry)
XX
XX Full length expanded cDNA sequence for human BAG-3 protein.
XX
XX Human BAG-3; Bcl-2 associated athanogene-3; apoptosis; cell migration;
KW tumour cell metastasis; Hsc70/Hsp70-regulating protein; metastasis;
KW tumour cell proliferation; steroid hormone receptor function; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 307..2034
FT /*tag= a
FT /product= "Human BAG-3 protein"
XX
XX WO200014106-A1.
XX
XX 16-MAR-2000.
XX
XX 09-SEP-1999; 99WO-US21053.
XX
XX 09-SEP-1998; 98US-0150489.
XX
XX (BURN-) BURNHAM INST.
PA

XX Reed JC, Takayama S;
XX WPI: 2000-256937/22.
DR P-PSDB: AAV70515.
XX BAG-1 related proteins from humans, *Caenorhabditis elegans* and
XX *Schistosoma* pome useful for modulating tumor cell
XX proliferation, cell migration and metastasis and steroid hormone
XX receptor function -
PS Claim 10; Fig 15; 132pp; English.
XX The present expanded cDNA sequence encodes human BAG-3 (Bcl-2 associated
XX athanogene-3) protein. BAG is a Hsc70/Hsp70-regulating protein
XX (Hsc70/Hsp70 is a molecular chaperone that participates in controlling
XX protein bioactivity, degradation; complex assembly/disassembly and
XX translocation across membranes). It competes with Hip for binding to the
XX Hsc70/Hsp70 ATPase binding domain and promotes substrate release. Gene
XX transfection studies indicate that BAG proteins influence a wide variety
XX of cellular phenotypes through their interactions with Hsc70/Hsp70.
XX including increasing resistance to apoptosis, promoting cell
XX proliferation, enhancing tumour cell migration and metastasis and
XX altering transcriptional activity of steroid hormones. The ability of
XX BAG-2 and BAG-3 proteins to interact in cells with Hsc70 was tested by
XX expressing these proteins with N-terminal flag group tags in 293T human
XX epithelial cells using immunoprecipitation assays. BAG also stimulates
XX Hsc70-mediated adenosine triphosphate (ATP) hydrolysis by accelerating
XX ADP/ATP exchange.
SO Sequence 2534 BP; 596 A; 806 C; 653 G; 479 T; 0 other;

Query Match 1.5%; Score 19; DB 21; Length 2534;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGCGCGCGCAG 76
DB 766 GTGGCAGCGCGCGCGCAG 784

RESULT 30
AAV84294
ID AAV84294 standard; DNA; 3636 BP.
AC AAV84294;
XX 12-APR-1999 (first entry)
DT XX
DE Rat (Na,K)-ATPase alpha 1 subunit (L879C mutant) DNA.
XX
KM Selectable marker; (Na,K)-ATPase; rat; gene therapy; mutant; ss.
XX
OS Rattus rattus.
OS Synthetic.
XX
PN WO9855603-A1.
XX
PD 10-DEC-1998.
XX
PF 04-JUN-1998; 98WO-SE01062.
XX
PR 04-JUN-1997; 97SE-0002120.
XX
PA (KARO-) KAROLINSKA INNOVATIONS AB.
XX
PI Belusa R;
XX
XX WPI: 1999-059908/05.
DR P-PSDB: AAV88287.
XX
PT New mutated (Na,K)-ATPase - useful as a selection marker

XX Disclosure; Page 45-49; 72pp; English.
PS This nucleotide sequence encodes a new mutated (Na,K)-ATPase alpha
XX 1 subunit (see AAV88287), in which the Leu-881 residue of the native
XX rat protein (see AAV88290) has been mutated to cysteine. This
XX mutated protein can be used as a novel selectable marker. The
XX mutation, obtained by mutagenesis of the (Na,K)-ATPase gene (see
XX AAV84297), causes a loss in sensitivity of the encoded enzyme to
XX ouabain. A claimed method for detecting transfection or
XX transformation of a eukaryotic cell comprises: (a) transforming or
XX transfecting the cell with a vector carrying a nucleic acid encoding
XX a mutated (Na,K)-ATPase gene of the invention, plus a gene of
XX interest; (b) incubating the cells with a cardioglycoside such
XX as digitoxin, digitoxigenin, digoxin, bufalin or (especially) ouabain;
XX (c) contacting the incubated cells with a fibrinogen-coated solid
XX matrix; and (d) detecting any cells bound to the matrix. The new
XX selectable marker is particularly useful where antibiotic resistance
XX selection is undesirable, e.g. in gene therapy.

Query Match 1.5%; Score 19; DB 20; Length 3636;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGCGCGCGCAG 76
DB 28 GTGGCAGCGCGCGCGCAG 46

RESULT 31
AAV84295
ID AAV84295 standard; DNA; 3636 BP.
AC AAV84295;
XX 12-APR-1999 (first entry)
DT XX
DE Rat (Na,K)-ATPase alpha 1 subunit (L881C mutant) DNA.
XX
KM Selectable marker; (Na,K)-ATPase; rat; gene therapy; ss.
XX
OS Rattus rattus.
OS Synthetic.
XX
PN WO9855603-A1.
XX
PD 10-DEC-1998.
XX
PF 04-JUN-1998; 98WO-SE01062.
XX
PR 04-JUN-1997; 97SE-0002120.
XX
PA (KARO-) KAROLINSKA INNOVATIONS AB.
XX
PI Belusa R;
XX
XX WPI: 1999-059908/05.
DR P-PSDB: AAV88288.
XX
PT New mutated (Na,K)-ATPase - useful as a selection marker
XX
PS Disclosure; Page 49-53; 72pp; English.
XX This nucleotide sequence encodes a new mutated (Na,K)-ATPase alpha
XX 1 subunit (see AAV88288), in which the Leu-881 residue of the native
XX rat protein (see AAV88290) has been mutated to cysteine. This
XX mutated protein can be used as a novel selectable marker. The
XX mutation, obtained by mutagenesis of the (Na,K)-ATPase gene (see
XX AAV84297), causes a loss in sensitivity of the encoded enzyme to
XX ouabain. A claimed method for detecting transfection or

transformation of a eukaryotic cell comprises: (a) transforming or transfecting the cell with a vector carrying a nucleic acid encoding a mutated (Na,K)-ATPase gene of the invention plus a gene of interest; (b) incubating the cells with a cardioglycoside such as digitoxin, digitogenin, digoxin, bufalin or (especially) ouabain; (c) contacting the incubated cells with a fibrinogen-coated solid matrix; and (d) detecting any cells bound to the matrix. The new selectable marker is particularly useful where antibiotic resistance selection is undesirable, e.g. in gene therapy.

Sequence 3636 BP; 850 A; 967 C; 995 G; 824 T; 0 other;

Query Match 1.5%; Score 19; DB 20; Length 3636;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GTGGCAGCGCGCGCGCAG 76
|||||
DB 28 gtgcagcgcgcgcgcgcag 46

RESULT 32

AAV84296
ID AAV84296 standard; DNA; 3636 BP.

AC AAV84296;

DT 12-APR-1999 (first entry)

DE Rat (Na,K)-ATPase alpha 1 subunit (L879C, L881C, mutant) DNA.

KW Selectable marker; (Na,K)-ATPase; rat; gene therapy; ss.

OS Rattus rattus.

OS Synthetic.

PN WO9855603-A1.

PD 10-DEC-1998.

PF 04-JUN-1998; 98WO-SE01062.

PR 04-JUN-1997; 97SE-0002120.

PA (KARO-) KAROLINSKA INNOVATIONS AB.

PI Belusa R;

DR WPI: 1999-059908/05.

P-PSDB; AAM88289.

PT New mutated (Na,K)-ATPase - useful as a selection marker

Disclosure; Page 54-58; 72pp; English.

This nucleotide sequence encodes a new mutated (Na,K)-ATPase alpha 1 subunit (see AAM88288), in which the Leu-879 and Leu-881 residues of the native rat protein (see AAM88290) have been mutated to cysteine. This mutated protein can be used as a novel selectable marker. The mutation, obtained by mutagenesis of the (Na,K)-ATPase gene (see AAV84297), causes a loss in sensitivity of the encoded enzyme to ouabain. A claimed method for detecting transfection or transformation of a eukaryotic cell comprises: (a) transforming or transfecting the cell with a vector carrying a nucleic acid encoding a mutated (Na,K)-ATPase gene of the invention, plus a gene of interest; (b) incubating the cells with a cardioglycoside such as digitoxin, digitogenin, digoxin, bufalin or (especially) ouabain; (c) contacting the incubated cells with a fibrinogen-coated solid matrix; and (d) detecting any cells bound to the matrix. The new selectable marker is particularly useful where antibiotic resistance selection is undesirable, e.g. in gene therapy.

SQ Sequence 3636 BP; 850 A; 966 C; 995 G; 825 T; 0 other;

Query Match 1.5%; Score 19; DB 20; Length 3636;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GTGGCAGCGCGCGCGCAG 76
|||||
DB 28 gtgcagcgcgcgcgcgcag 46

RESULT 33

AAV84297
ID AAV84297 standard; DNA; 3636 BP.

AC AAV84297;

DT 12-APR-1999 (first entry)

DE Rat (Na,K)-ATPase alpha 1 subunit gene coding region.

KW Selectable marker; (Na,K)-ATPase; rat; gene therapy; ss.

OS Rattus rattus.

PN WO9855603-A1.

PD 10-DEC-1998.

PF 04-JUN-1998; 98WO-SE01062.

PR 04-JUN-1997; 97SE-0002120.

PA (KARO-) KAROLINSKA INNOVATIONS AB.

PI Belusa R;

DR WPI: 1999-059908/05.

P-PSDB; AAM88290.

PT New mutated (Na,K)-ATPase - useful as a selection marker

Disclosure; Page 59-63; 72pp; English.

This nucleotide sequence encodes the rat (Na,K)-ATPase alpha 1 subunit (see AAM88290). (Na,K)-ATPase is an ion transporter essential to the cell. The invention provides new, mutated (Na,K)-ATPases (see AAM88287-89) that can be used as novel selectable markers. Appropriate mutation of the (Na,K)-ATPase gene (see AAV84294-96) causes a loss in sensitivity of the encoded enzyme to ouabain. A claimed method for detecting transfection or transformation of a eukaryotic cell comprises: (a) transforming or transfecting the cell with a vector carrying a nucleic acid encoding a mutated (Na,K)-ATPase gene of the invention, plus a gene of interest; (b) incubating the cells with a cardioglycoside such as digitoxin, digitogenin, digoxin, bufalin or (especially) ouabain; (c) contacting the incubated cells with a fibrinogen-coated solid matrix; and (d) detecting any cells bound to the matrix. The new selectable markers are particularly useful where antibiotic resistance selection is undesirable, e.g. in gene therapy.

Sequence 3636 BP; 850 A; 968 C; 995 G; 823 T; 0 other;

Query Match 1.5%; Score 19; DB 20; Length 3636;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GTGGCAGCGCGCGCGCAG 76
|||||
DB 28 gtgcagcgcgcgcgcgcag 46

RESULT 34

AAAF44683 standard; cDNA: 7328 BP.

ID: AAAF44683

AC: AAAF44683

DE: 27-MAR-2001 (first entry)

XX: Novel protein kinase cDNA, SEQ ID NO: 63.

XX: Human; mouse; protein kinase; antihypertensive; osteoporosis; immunosuppressive; cardiogenic; renal; anti-inflammatory; antidiabetic; dermatological; antidiabetic; anti-infective; gene therapy; vaccine;

XX: immune disorder; cardiovascular disease; neurodegenerative disease; cancer; autoimmune disorder; stroke; inflammatory bowel disease; inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.

XX: Homo sapiens.

XX: MO200073469-A2.

XX: 07-DEC-2000.

XX: 26-MAY-2000. 2000MO-US14842.

XX: 28-MAY-1999. 99US-0136503.

XX: (SUGC-) SUGEN INC.

XX: Plowman GD, Martinez R, Whyte D, Sudersanam S;

XX: WPI: 2001-032161/04.

XX: P-PSDB: AAB65656.

XX: Nucleic acids encoding kinase polypeptides, useful for diagnosing and treating immune-related diseases and disorders, cardiovascular disease, neurodegenerative diseases and/or cancers -

XX: Example 1; Fig 2; 310pp; English.

XX: The present sequence encodes a novel protein kinase. The nucleic acids and the protein kinases they encode may be used in the treatment and diagnosis of diseases associated with inappropriate kinase expression, such as immune-related diseases and disorders, cardiovascular disease, neurodegenerative diseases and/or cancers. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays. The kinase polypeptides may be used as antigens in the production of antibodies of kinase expression and activity. Anti-kinase antibodies and kinase antagonists may also be used to down regulate kinase expression and activity. Diseases related to kinase expression and activity include rheumatoid arthritis, atherosclerosis, autoimmune disorders, complications of organ transplantation, myocardial infarction, immune disorders, cardiomyopathies, strokes, renal failure, oxidative-stress related disorders, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and reproductive disorders.

SQ: Sequence 7328 BP; 2036 A; 1850 C; 1655 G; 1785 T; 2 other;

Query Match 1.5%; Score 19; DB 22; Length 7328;

Best Local Similarity 100.0%; Pred. No. 17; Mismatches 0; Indels 0; Gaps 0;

QY: 58 GTGGCAGCGCGCGCGCAG 76

DB: 589 GTGGCAGCGCGCGCGCAG 607

RESULT 35

AAZ38983

ID: AAZ38983 standard; DNA: 30 BP.

AC: AAZ38983

DE: 28-FEB-2000 (first entry)

XX: Expression vector construction primer SEQ ID NO:18.

XX: 1,3-Propanediol; vitamin B12 transport; glycerol dehydratase; Btub;

XX: 1,3-Propanediol oxidoreductase; vitamin B12 receptor precursor; Btub;

XX: vitamin B12 transport system permease protein; Btub; polyester fibre;

XX: vitamin B12 transport ATP-binding protein; polyurethane; primer; ss.

XX: Synthetic.

XX: WO9958686-A2.

XX: 18-NOV-1999.

XX: 12-MAY-1999. 99MO-US10356.

XX: 12-MAY-1998. 98US-0085190.

XX: (DUPD) DU PONT DE NEMOURS & CO E I.

XX: (GENV) GENENCOR INT INC.

XX: Whited GM, Bulhuis B, Trimbur DE, Gatenby AA;

XX: WPI: 2000-053104/04.

XX: Production of a monomer which is used in the manufacture of polyester fiber, polyurethane and cyclic compounds -

XX: Example 1; Page 56; 61pp; English.

XX: The present invention describes the monomer, 1,3-propanediol (I), which is biologically produced by using a transformed host cell (C) comprising at least one copy of gene encoding Btub, Btuc and Btnd. (I) is produced by the steps comprising: (1) contacting the transformed (C) with at least one fermentable common source and an effective amount of vitamin B12. The one transformed (C) comprises at least one copy of genes encoding a protein having dehydratase activity (a), a protein having an oxidoreductase activity (b), a vitamin B12 receptor precursor protein (c), a vitamin B12 transport system permease protein (d), and a vitamin B12 transport ATP or GTP-binding protein (e); and (2) recovering (I) produced by (1). The method is used to produce (I) which is employed in the manufacture of polyester fibre, polyurethane and cyclic compounds. The present invention produces 1,3-propanediol rapidly and it is inexpensive. The present sequence represents a primer used in the construction of expression vectors in the exemplification of the present invention.

SQ: Sequence 30 BP; 8 A; 4 C; 6 G; 12 T; 0 other;

Query Match 1.4%; Score 18; DB 21; Length 30;

Best Local Similarity 100.0%; Pred. No. 49; Mismatches 0; Indels 0; Gaps 0;

QY: 872 ATGTGAGCTTACCA 889

DB: 4 attgtgagcttaccac 21

RESULT 36

AAV90298 standard; cDNA: 346 BP.

ID: AAV90298

AC: AAV90298

DE: 15-FEB-1999 (first entry)

XX: EST clone D1482.

KM Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KM tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KM receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KM gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO9845436-A2.
XX
PD 15-OCT-1998.
XX
PE 10-APR-1998; 98WO-US06955.
XX
PR 10-APR-1997; 97US-0838821.
XX
PA (GEMV) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-070077/06.
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
PS Claim 1: Page 496; 618pp; English.
XX
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
SQ Sequence 346 BP; 128 A; 62 C; 101 G; 55 T; 0 other;

Query Match 1.4%; Score 18; DB 20; Length 346;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 680 GGAACACGAAATGAG 697
|||||
DB 266 ggaacacgagaatgag 283

RESULT 37
AAC01530
ID AAC01530 standard; cDNA; 402 BP.
XX
AC AAC01530;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1528.
XX
KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.

XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
PA (GEMV) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
DR
DR P-PDB; AAC01524.
XX
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX
PS Claim 1; SEQ ID 1528; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 402 BP; 144 A; 68 C; 113 G; 70 T; 7 other;

Query Match 1.4%; Score 18; DB 21; Length 402;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 680 GGAACACGAAATGAG 697
|||||
DB 244 ggaacacgagaatgag 261

RESULT 38
AAF11454/C
ID AAF11454 standard; cDNA; 543 BP.
XX
AC AAF11454;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus niger EST SEQ ID NO:3977.
XX
KM Multiple gene expression; filamentous fungal cell; EST;
KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KM culture condition; environmental stress; spore morphogenesis;
KM metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus niger.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PE 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;

XX WPI: 2000-594572/56.

XX Monitoring differential expression of genes in filamentous fungal cells
PT Uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -

PS Claim 87; Page 1764; 3161pp; English.

CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random DNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AA07478 to AA11247 represents ESTs from
CC *Fusarium venenatum*; AA11248 to AA11853 represents ESTs from *Aspergillus*
CC *niger*; AA11854 to AA14878 represents ESTs from *Aspergillus oryzae*; and
CC AA14879 to AA15337 represents ESTs from *Trichoderma reesei*, which are
CC all specifically claimed in the present invention.

CC Sequence 543 BP; 128 A; 128 C; 121 G; 166 T; 0 other;

SO Query Match 1.4%; Score 18; DB 21; Length 543;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCAGG 77
DB 163 GGCAGCGCGCGCGCAGG 146

RESULT 39

AAH1335/C
ID AAH1335 standard; cDNA; 582 BP.

AC AAH1335;

DT 26-JUN-2001 (first entry)

DE Human cDNA clone (3'-primer); SEQ ID NO:9970.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

PS Claim 3; SEQ ID 9970; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complement and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

CC Sequence 582 BP; 142 A; 120 C; 83 G; 229 T; 8 other;

SO Query Match 1.4%; Score 18; DB 22; Length 582;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 870 CCATTGTGATGCTTTAC 887
DB 106 CCATTGTGATGCTTTAC 89

RESULT 40

AAH67523/C
ID AAH67523 standard; DNA; 588 BP.

AC AAH67523;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 2558.

KW Corynebacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

OS Corynebacterium glutamicum.

PN EP1108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX

PA (KYOW) KYOMA HAKKO KOCYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI: 2001-376931/40.
 DR P-PSDB: AAG92304.
 XX
 PT Novel polynucleotides derived from *Corynebacterium* bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 8; SEQ ID NO: 2558; 246bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the *Corynebacterium* bacterium *Corynebacterium glutamicum*. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of *Corynebacterium* bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from *Corynebacterium* bacterium, and identifying a homologue of a gene derived
 CC from *Corynebacterium* bacterium. *Corynebacterium* bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 XX
 SQ Sequence 588 BP; 115 A; 134 C; 186 G; 153 T; 0 other:
 XX

Query Match 1.4%; Score 18; DB 22; Length 588;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 993 CATCGTGTTCATCTGCAT 1010
 ||||||||||||||||
 Db 91 CATCGTGTTCATCTGCAT 74

RESULT 41
 AAF11560/c
 ID AAF11560 standard; CDNA: 593 BP.
 XX
 AC AAF11560;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE *Aspergillus niger* EST SEQ ID NO:4083.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; *Fusarium venenatum*; *Aspergillus niger*;
 KW *Aspergillus oryzae*; *Trichoderma reesei*; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS *Aspergillus niger*.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PE 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX
 DR WPI: 2000-594572/56.

XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 87; Page 1796; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from
 CC *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus*
 CC *niger*; AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
 CC all specifically claimed in the present invention.
 CC
 XX
 SQ Sequence 593 BP; 104 A; 198 C; 160 G; 124 T; 7 other:
 XX

Query Match 1.4%; Score 18; DB 21; Length 593;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 TGGCAGCGCGCGGCGCAG 76
 ||||||||||||||||
 Db 286 TGGCAGCGCGCGCGCAG 269

RESULT 42
 AAT37915/c
 ID AAT37915 standard; CDNA: 624 BP.
 XX
 AC AAT37915;
 XX
 DT 28-APR-1997 (first entry)
 XX
 DE VEGF-B186 coding sequence.
 XX
 KW Endothelial cell; proliferation; vascular endothelial growth factor; VEGF;
 KW VEGF; endothelium; mesodermal cell; cationic dimer; tissue regeneration;
 KW vascular permeability factor; cell mitogen; angiogenesis; cell growth;
 KW embryonic development; wound healing; tissue reorganisation; antibody;
 KW cancer; metastatic risk; tumour cell; mouse; ss.
 XX
 OS *Homo sapiens*.
 XX
 PN WO9626736-A1.
 XX
 PD 06-SEP-1996.
 XX
 PE 01-MAR-1996; 96WO-US02957.
 XX
 PR 06-DEC-1995; 95US-0569063.
 PR 01-MAR-1995; 95US-0397651.
 PR 06-JUN-1995; 95US-0469427.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.

(UYHE-) UNIV HELSINKI LICENSING LTD OY.
 Alitalo K, Eriksson U, Olofsson B, Pajusola K;
 WPI: 1996-412582/41.
 P-PSDB: AAM04831.
 Vascular endothelial growth factor VEGF-B proteins - useful to accelerate angiogenesis in wound healing, also related nucleic acid and antibodies for cancer diagnosis
 Claim 1: Page 61-62; 107pp; English.
 AA17909-17915 represent the coding sequences for the vascular endothelial growth factor (VEGF) proteins of the invention, which promote endothelial or mesodermal-cell proliferation. VEGF is also a glycosylated cationic dimer, and is sometimes referred to as vascular permeability factor (VPF). VEGF has diverse effects, depending on the specific biological context in which it is found. VEGF is a potent endothelial cell mitogen, and directly contributes to induction of angiogenesis in vivo by promoting endothelial cell growth during normal angiogenic development, wound healing, and tissue regeneration/reorganisation. The VEGF proteins of the invention share the angiogenic and other properties of VEGF, but are distributed and expressed in tissues differently to VEGF. The proteins can therefore be used to accelerate angiogenesis in wound healing. Antibodies against the proteins can be used for inhibiting angiogenesis. The antibodies can also be used diagnostically to quantitatively detect the invention of complementary to the coding sequences for the proteins of the invention of can also be used to detect VEGF-B coding sequences. Quantification of VEGF-B in cancer biopsy specimens may be useful as an indicator of metastatic risk. VEGF-B expression in a cell can be retarded using antisense sequences direct against the VEGF coding sequences, this is especially useful in retarding VEGF expression in tumour cells.
 Sequence 624 BP; 106 A; 235 C; 179 G; 104 T; 0 other;
 SQ
 Query Match 1.4%; Score 18; DB 17; Length 624;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 60 GGCAGCGCGCGCGCAGG 77
 DB 585 GGCAGCGCGCGCGCAGG 568
 RESULT 43
 AAV63569/c
 ID AAV63569 standard; cDNA: 624 BP.
 AC AAV63569;
 DT 29-JAN-1999 (first entry)
 DE cDNA clone encoding vascular endothelial growth factor (VEGF)-B186.
 KW Vascular endothelial growth factor; VEGF; proliferation; human; endothelial cell; angiogenesis; tissue growth; organ repair; ss.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT 1..624
 FT CDS /*tag= a
 FT /product= VEGF-B186
 PN US5840693-A.
 PD 24-NOV-1998.
 PF 01-MAR-1996; 96US-0609443.
 XX

01-MAR-1996; 96US-0609443.
 01-MAR-1995; 95US-0397651.
 06-JUN-1995; 95US-0469427.
 06-DEC-1995; 95US-0569063.
 (LUDW-) LUDWIG INST CANCER RES.
 (UYHE-) UNIV HELSINKI LICENSING LTD OY.
 Alitalo K, Eriksson U, Olofsson B, Pajusola K;
 WPI: 1999-034079/03.
 P-PSDB: AAM80495.
 Vascular endothelial growth factor-B isoforms and DNA encoding them useful for inducing angiogenesis and cellular proliferation, and raising antibodies to inhibit activities in e.g. tumours
 Claim 22; Fig 14; 52pp; English.
 The present sequence encodes human vascular endothelial growth factor (VEGF)-B186. VEGF proteins are used for promoting proliferation of endothelial cells and for stimulating angiogenesis (the proliferation of new capillaries form pre-existing blood vessels). These activities are useful for treating tissue growth and repair, including organ repair. This is also useful in pregnancy, in follicle development, as these processes must occur in development of the placenta. The proteins can also be used to raise antibodies, either for use in detection of the proteins or as inhibitors of their action. This is especially useful as angiogenesis is required by tumours as they need new blood supplies to grow and proliferate.
 Sequence 624 BP; 106 A; 235 C; 179 G; 104 T; 0 other;
 SQ
 Query Match 1.4%; Score 18; DB 20; Length 624;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 60 GGCAGCGCGCGCGCAGG 77
 DB 585 GGCAGCGCGCGCGCAGG 568
 RESULT 44
 AAA91005/c
 ID AAA91005 standard; DNA: 624 BP.
 AC AAA91005;
 DT 05-APR-2001 (first entry)
 DE Human VEGF-3 coding sequence.
 KW Human; angiogenic protein; wound healing; vascular tissue repair; peripheral arterial disease; critical limb ischaemia; coronary disease; angiogenesis; tumour; inflammation; diabetic retinopathy; psoriasis; rheumatoid arthritis; autoimmune disease; allergy; cancer; therapy; infectious disease; neurodegeneration;
 KW Vascular endothelial growth factor-3; VEGF-3; ss.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT 1..624
 FT CDS /*tag= a
 FT /product= VEGF-3
 PN WO200075163-A1.
 PD 14-DEC-2000.
 PF 01-JUN-2000; 2000WO-US14925.
 XX

PR 03-JUN-1999; 9905-0137796.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Hu J, Cao L;
XX WPI: 2001-071057/08.
DR P-PSDB: AAY97571.
XX
PT New nucleic acid encoding angiogenic proteins, useful e.g. for
PT promoting healing of wounds and treating peripheral arterial disease,
PT critical limb ischemia or coronary disease
XX
PS Claim 1: Page 224; 244pp; English.
XX
CC This sequence encodes vascular endothelial growth factor-3 (VEGF-3),
CC which is an angiogenic protein of the invention. The angiogenic proteins
CC and the DNA sequences encoding them, are used to prevent, treat or
CC ameliorate disease and to detect diseases, or susceptibility, by
CC detecting mutations or the presence or amount of angiogenic protein
CC expression. Particularly they are used to stimulate wound healing,
CC growth of damaged bone and tissue, and for repair of vascular tissue,
CC especially peripheral arterial disease, critical limb ischemia or
CC coronary disease. Antagonists of the sequences are used to inhibit
CC angiogenesis in tumours and to treat inflammation (where associated with
CC increased vascular permeability), diabetic retinopathy, rheumatoid
CC arthritis or psoriasis. Agonists are also useful for stimulating
CC (lymph)angiogenesis. The proteins are also used to identify specific
CC binding agents (potential therapeutic agents) and to raise antibodies.
CC The antibodies are useful as therapeutic (antagonists); for detection,
CC purification and targeting of proteins for in vivo or in vitro diagnosis
CC (including imaging) or for therapy (including when linked to e.g. a label
CC or cytotoxic); and for immunotyping of cells, e.g. for detecting minimal
CC residual disease or hematopoietic progenitor/stem cells. It is also
CC contemplated that the sequences might be useful for treating a very wide
CC range of other disorders, e.g. autoimmune diseases; allergy; cancer;
CC infectious diseases (viral, bacterial, fungal or parasitic);
CC neurodegeneration, also as chemotactic agents or for stimulating
CC regeneration of the nervous system etc.
XX
SO Sequence 624 BP; 106 A; 235 C; 179 G; 104 T; 0 other:
XX
Query Match 1.4%; Score 18; DB 22; Length 624;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 60 GGCAGCGCGCGCGCAGG 77
DB 585 GGCAGCGCGCGCGCAGG 568
XX
RESULT 45
ID AAA68125 standard; DNA; 762 BP.
XX
AC AAA68125;
XX
DT 24-OCT-2000 (first entry)
XX
DE Pinus radiata flavonol glucosyl transferase DNA sequence SEQ ID NO:218.
XX
XX Plant; lignin; lignin biosynthetic pathway; Eucalyptus grandis;
KM Pinus radiata; Monterey pine; ds.
XX
OS Pinus radiata.
XX
PN WO200022099-A1.
XX
PD 20-APR-2000.
XX
PF 06-OCT-1999; 99WO-N200168.
XX

PR 09-OCT-1998; 9805-0169789.
PR 14-JUL-1999; 9905-0143811.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN, Havukkala JU;
XX
DR WPI: 2000-317962/27.
XX
PT Novel polynucleotide encoding enzymes involved in lignin-biosynthetic
PT pathway useful for producing transgenic plants especially eucalyptus
PT and pine species having altered lignin content, composition and
PT structure
XX
PS Claim 1: Page 124; 213pp; English.
XX
CC The present invention describes isolated polynucleotides and proteins
CC encoding and representing the enzymes cinnamate 4-hydroxylase (C4H),
CC coumarate 3-hydroxylase (C3H), phenolase (PNL), O-methyl transferase
CC (OMT), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl-CoA reductase
CC (CCR), phenylalanine ammonia-lyase (PAL), 4-coumarate:CoA ligase (4CL),
CC coniferol glucosyl transferase (CGT), coniferin beta-glucosidase (CBG),
CC laccase, peroxidase, ferulate-5-hydroxylase (F5H), alpha-amylase,
CC caffeic acid methyl transferase, caffeoyl CoA methyl transferase,
CC coumarate CoA ligase, cytochrome P450 1X11, diphenol oxidase, flavanol
CC glucosyl transferase, flavonoid hydroxylase, and isoflavone reductase,
CC which are involved in the lignin biosynthetic pathway. The
CC polynucleotides can be used for modulating lignin content, lignin
CC composition and the structure of a plant, especially eucalyptus and pine
CC species, and for modifying the activity of an enzyme involved in lignin
CC biosynthetic pathway, and for producing a plant having altered lignin
CC content, composition and structure. They can be used for designing probes
CC and primers useful for detecting similar DNA and RNA sequences in any
CC organism and for PCR amplification. The lignin content can be efficiently
CC modified using the polynucleotides. AA67908 to AAA68201 and AAB16341 to
CC AAB16449 represent polynucleotide and protein sequences used in the
CC exemplification of the present invention.
XX
SO Sequence 762 BP; 227 A; 131 C; 202 G; 202 T; 0 other:
XX
Query Match 1.4%; Score 18; DB 21; Length 762;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 764 CATGGAAGCATAGTG 781
DB 241 catggaagcattagtg 258
XX

Search completed: March 29, 2002, 20:56:33
Job time: 2153 sec

Mon Apr 1 08:00:08 2002

us-09-415-540-2.oli.rng

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 19:55:10 ; Search time 90.84 Seconds
(without alignments)
3178.770 Million cell updates/sec

Title: US-09-415-540-2

Perfect score: 1275

Sequence: 1 CACAGAGTTGGGCTCTCT.....GAATTAATTTTGTCTG 1275

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 11328999 residues

Word size: 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/1/lna/PTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/1/lna/backfillseq.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1275	100.0	1275	2	US-08-741-437-2
2	1275	100.0	1275	2	US-09-134-593-2
3	18	1.4	624	2	US-08-609-443B-14
4	18	1.4	624	2	US-08-569-063C-14
5	18	1.4	4825	6	5459251-1
6	18	1.4	5086	2	US-08-465-485A-19
7	18	1.4	5086	2	US-08-365-486A-14
8	18	1.4	5086	3	US-09-080-285-19
9	18	1.4	5086	4	US-08-880-342-14
10	18	1.4	5086	5	PCT-US93-05651-4
11	18	1.4	5086	5	PCT-US93-06251-2
12	18	1.4	5104	6	5506344-1
13	17	1.3	77	2	US-08-180-524-5
14	17	1.3	77	2	US-08-975-166-5
15	17	1.3	146	2	US-08-180-524-3
16	17	1.3	146	2	US-08-975-166-3
17	17	1.3	461	1	US-08-454-720A-37
18	17	1.3	617	4	US-09-328-111-577
19	17	1.3	718	3	US-09-154-083-31
20	17	1.3	810	3	US-09-097-072-1
21	17	1.3	900	3	US-08-886-638-1
22	17	1.3	900	3	US-09-097-072-3
23	17	1.3	900	4	US-09-212-988-1
24	17	1.3	1068	5	PCT-US91-00899-13
25	17	1.3	1215	5	PCT-US91-00899-5
26	17	1.3	1488	1	US-07-914-281-9
27	17	1.3	1488	1	US-08-393-246-9

28	17	1.3	1488	1	US-08-525-058A-9	Sequence 9, Appl1
29	17	1.3	1488	2	US-08-696-731-9	Sequence 9, Appl1
30	17	1.3	1488	4	US-09-042-531-9	Sequence 9, Appl1
31	17	1.3	1846	2	US-08-365-486A-16	Sequence 16, Appl1
32	17	1.3	1846	4	US-08-880-342-16	Sequence 16, Appl1
33	17	1.3	2134	5	PCT-US96-06427-3	Sequence 3, Appl1
34	17	1.3	2134	5	PCT-US96-06427-3	Sequence 3, Appl1
35	17	1.3	2175	4	US-08-482-073-10	Sequence 10, Appl1
36	17	1.3	2861	4	US-08-006-676B-2	Sequence 2, Appl1
37	17	1.3	3319	1	US-08-282-845-1	Sequence 1, Appl1
38	17	1.3	3319	2	US-08-428-414A-4	Sequence 4, Appl1
39	17	1.3	3319	5	PCT-US94-00324-2	Sequence 1, Appl1
40	17	1.3	3467	3	US-08-972-719-1	Sequence 1, Appl1
41	17	1.3	3467	4	US-08-753-038-1	Sequence 7, Appl1
42	17	1.3	3647	1	US-07-914-281-7	Sequence 7, Appl1
43	17	1.3	3647	1	US-08-393-246-7	Sequence 7, Appl1
44	17	1.3	3647	1	US-08-525-058A-7	Sequence 7, Appl1
45	17	1.3	3647	1	US-08-525-058A-7	Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-08-741-437-2
Sequence 2, Application US/08741437
Patent No. 5843665
GENERAL INFORMATION:
APPLICANT: Hawkins, Phillip R.
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/741,437
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0148 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 845-0555
TELEFAX: (415) 845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-08-741-437-2

Query Match 100.0%; Score 1275; DB 2; Length 1275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 CAAGAGTTTNGGGGCTCTCTCTTGTGAGTGGGCGCCGCTGGGGCTGTGGCTCTGTG 60
DB 1 CAAGAGTTTNGGGGCTCTCTCTTGTGAGTGGGCGCCGCTGGGGCTGTGGCTCTGTG 60
OY 61 GCAGGGGGGGGGGAGGAGCTCCGGCATTATAGCGGGCTTCAGACCGAGAGAGCGCGCGC 120
DB 61 GCAGGGGGGGGGGAGGAGCTCCGGCATTATAGCGGGCTTCAGACCGAGAGAGCGCGCGC 120
OY 121 GCCCTTCCCTGGAGTACCGAGTCTCTCTCAAAATGAGAAAGACATATATCTCC 180
DB 121 GCCCTTCCCTGGAGTACCGAGTCTCTCTCAAAATGAGAAAGACATATATCTCC 180
OY 181 ATTTCATATATTCATTTATGAGATAGAGATGTGTTACATGATGATGAGTACC 240
DB 181 ATTTCATATATTCATTTATGAGATAGAGATGTGTTACATGATGATGAGTACC 240
OY 241 ACGTGTCTATATGCAAAATGAGATGTCTACAAAGAGCCCTTTAAACCTATTAAACA 300
DB 241 ACGTGTCTATATGCAAAATGAGATGTCTACAAAGAGCCCTTTAAACCTATTAAACA 300
OY 301 AGATGTGAAAAAGAAAACTTCTGCTATGTGCGAATTTGTTCCCTATTAAGATATAT 360
DB 301 AGATGTGAAAAAGAAAACTTCTGCTATGTGCGAATTTGTTCCCTATTAAGATATAT 360
OY 361 CTGGAACATATGTGCTCCCTCAGACCTTGGGAAGCCAGGGGACATGATTAACATAC 420
DB 361 CTGGAACATATGTGCTCCCTCAGACCTTGGGAAGCCAGGGGACATGATTAACATAC 420
OY 421 TGGCTGTGTGTGCAATGACCAATTGATGTGTGTAATTTGAGAGCAAGATGTGC 480
DB 421 TGGCTGTGTGTGCAATGACCAATTGATGTGTGTAATTTGAGAGCAAGATGTGC 480
OY 481 AAGAGTGAATAATTTGGGTGAAGTCTAGCATATTTGGCTATGATTTGACGAGGGGA 540
DB 481 AAGAGTGAATAATTTGGGTGAAGTCTAGCATATTTGGCTATGATTTGACGAGGGGA 540
OY 541 AACCGACTGGAAGTCAATTTGCTTAATGTGATGATCCTGATGACCAATTTAATGA 600
DB 541 AACCGACTGGAAGTCAATTTGCTTAATGTGATGATCCTGATGACCAATTTAATGA 600
OY 601 TATCATATGATGTCAAAACGGCTGAACCTGGCTACTTAGAGCTACTGTGGCTTTAG 660
DB 601 TATCATATGATGTCAAAACGGCTGAACCTGGCTACTTAGAGCTACTGTGGCTTTAG 660
OY 661 AAGGTATAGTCTCTGATGGAACCAAGAAATAGTTGGTGTATGAGCAATTTAA 720
DB 661 AAGGTATAGTCTCTGATGGAACCAAGAAATAGTTGGTGTATGAGCAATTTAA 720
OY 721 AGATTAAGACTTTGCCATTTGATTTATTTAAAGCACTCATGACCAATTTGAAGCAATTTAGT 780
DB 721 AGATTAAGACTTTGCCATTTGATTTATTTAAAGCACTCATGACCAATTTGAAGCAATTTAGT 780
OY 781 GACTAAGAAAAAGAAATGAAATGCACTGTCATGATATACACTTTGTCTGAGAGCCC 840
DB 781 GACTAAGAAAAAGAAATGAAATGCACTGTCATGATATACACTTTGTCTGAGAGCCC 840
OY 841 CTTCAAGTGTGATCCGATGCTGCGAGAGCAATGTGATGCTTTTACCAACCACTGTGA 900
DB 841 CTTCAAGTGTGATCCGATGCTGCGAGAGCAATGTGATGCTTTTACCAACCACTGTGA 900
OY 901 ATCTGCTGCACAGTACCAAGAGAGCTGATAGTGTTCATCACCAGAAAAACTATG 960
DB 901 ATCTGCTGCACAGTACCAAGAGAGCTGATAGTGTTCATCACCAGAAAAACTATG 960
OY 961 AATATTTCTGGAATCAAGCTGATATGCTATCATCTGTTATCTGGATGATTAAGAAG 1020
DB 961 AATATTTCTGGAATCAAGCTGATATGCTATCATCTGTTATCTGGATGATTAAGAAG 1020
OY 1021 TAAAGTAGTACTTTCAAGCTTTAATTTGATGAACATCTTAACATAAGTAATTC 1080
DB 1021 TAAAGTAGTACTTTCAAGCTTTAATTTGATGAACATCTTAACATAAGTAATTC 1080
```

```
OY 1081 TGTGTGACTATATCAATATATACAGAAATGTTATCACTTAAGCAATTTTCAATCTCA 1140
DB 1081 TGTGTGACTATATCAATATATACAGAAATGTTATCACTTAAGCAATTTTCAATCTCA 1140
OY 1141 ACTAAGATATATTTTACACATGCTTAAATATCAACAGATGTGATTTGGAAGTCACTT 1200
DB 1141 ACTAAGATATATTTTACACATGCTTAAATATCAACAGATGTGATTTGGAAGTCACTT 1200
OY 1201 GTGAATAGATGTGCAAGGGGAGACATATTTGATGTATGTTACATATTTAGAAAT 1260
DB 1201 GTGAATAGATGTGCAAGGGGAGACATATTTGATGTATGTTACATATTTAGAAAT 1260
OY 1261 AAAATATTTTGTG 1275
DB 1261 AAAATATTTTGTG 1275

RESULT 2
US-09-134-593-2
: Sequence 2, Application US/09134593
: Patent No. 5981232
: GENERAL INFORMATION:
: APPLICANT: Hawkins, Phillip R.
: APPLICANT: Hillman, Jennifer L.
: TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: California
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/134,593
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/741,437
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0148 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 855-0555
: TELEFAX: (415) 845-4166
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1275 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: IMMEDIATE SOURCE:
: LIBRARY:
: CLONE: Consensus
: US-09-134-593-2

Query Match 100.0%; Score 1275; DB 2; Length 1275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAGAGTTTNGGGGCTCTCTCTTGTGAGTGGGCGCCGCTGGGGCTGTGGCTCTGTG 60
DB 1 CAAGAGTTTNGGGGCTCTCTCTTGTGAGTGGGCGCCGCTGGGGCTGTGGCTCTGTG 60
OY 61 GCAGGGGGGGGGGAGGAGCTCCGGCATTATAGCGGGCTTCAGACCGAGAGAGCGCGCGC 120
```

Db 61 GCACGGCGGGGCGAGAGCTCCGACTATGAGCGCTTCAGCAGCCAGGAGCGCGCGC 120
OY 121 GGCCTTCCCTGGAGTACCGAGTCTCCCTCAAAAATGAGAAAGACAAATATATCC 180
Db 121 GGCCTTCCCTGGAGTACCGAGTCTCCCTCAAAAATGAGAAAGACAAATATATCC 180
OY 181 ATTTTCATGATATTCATTTATGAGATAGAGATGTGTTTCATGAGTATGAGATACC 240
Db 181 ATTTTCATGATATTCATTTATGAGATAGAGATGTGTTTCATGAGTATGAGATACC 240
OY 241 ACCTGCTCTATATCAAAAATGAGATGTGTTTCATGAGTATGAGATACC 300
Db 241 ACCTGCTCTATATCAAAAATGAGATGTGTTTCATGAGTATGAGATACC 300
OY 301 AGATGTGAAAAAAGAAAAACCTCCCTATGTGTCGAATTTGTTCCCGTATAAGATATAT 360
Db 301 AGATGTGAAAAAAGAAAAACCTCCCTATGTGTCGAATTTGTTCCCGTATAAGATATAT 360
OY 361 CTGGAATATGTGTCGCAATCCCTCAGACTGGGAGACCCAGGACCAATGATATAACATAC 420
Db 361 CTGGAATATGTGTCGCAATCCCTCAGACTGGGAGACCCAGGACCAATGATATAACATAC 420
OY 421 TGCGTGTGTGTCGCAATGACCAATGTGTCGAATTTGTTCCCGTATAAGATATATGTC 480
Db 421 TGCGTGTGTGTCGCAATGACCAATGTGTCGAATTTGTTCCCGTATAAGATATATGTC 480
OY 481 AAGAGTGAATTAATTTGGCGGAGAAATTCAGCATATTTGGCTATGATGAGAGAGGGA 540
Db 481 AAGAGTGAATTAATTTGGCGGAGAAATTCAGCATATTTGGCTATGATGAGAGAGGGA 540
OY 541 AACCGACTGGAAGTCAATTTGCCATTTAATGTGATGATCCCTGATCAGCCAAATTAATGA 600
Db 541 AACCGACTGGAAGTCAATTTGCCATTTAATGTGATGATCCCTGATCAGCCAAATTAATGA 600
OY 601 TATCAATGATGTCAAAAGGCTGAAACCTGGCTACTAGAACCTGCTGGAAGTGTAG 660
Db 601 TATCAATGATGTCAAAAGGCTGAAACCTGGCTACTAGAACCTGCTGGAAGTGTAG 660
OY 661 AAGGTATAGGTTCCTGTCGTAAGAAACCAAGAAATGAGTTCGTTTATGAGAAATTA 720
Db 661 AAGGTATAGGTTCCTGTCGTAAGAAACCAAGAAATGAGTTCGTTTATGAGAAATTA 720
OY 721 AGATAGGACTTTGCCATTTGATTTAATAAGCACTCATGACATTGGAAGCATTTAGT 780
Db 721 AGATAGGACTTTGCCATTTGATTTAATAAGCACTCATGACATTGGAAGCATTTAGT 780
OY 781 GACTAAGAAAAAGAAAGAAAGCAATGATGATGATCAATCAACTTTGCTGAGAGCCC 840
Db 781 GACTAAGAAAAAGAAAGAAAGCAATGATGATGATCAATCAACTTTGCTGAGAGCCC 840
OY 841 CTTCAAGTGTATCTGATGCTGTCGACAGAGCCATTTGATGCTTTACCAACCCCTGTA 900
Db 841 CTTCAAGTGTATCTGATGCTGTCGACAGAGCCATTTGATGCTTTACCAACCCCTGTA 900
OY 901 ATCTGCTGACAGTATACCAAGAGTGTATGATGATGATGATGATGATGATGATGATG 960
Db 901 ATCTGCTGACAGTATACCAAGAGTGTATGATGATGATGATGATGATGATGATGATG 960
OY 961 AGATTTCTCTGGAATACAGCTGATGATGATGATGATGATGATGATGATGATGATG 1020
Db 961 AGATTTCTCTGGAATACAGCTGATGATGATGATGATGATGATGATGATGATGATG 1020
OY 1021 TAAAGTAGTATGCTTTCAAGCTTTAATTTGTAAGAACTCATTAAGTAAGTAATTC 1080
Db 1021 TAAAGTAGTATGCTTTCAAGCTTTAATTTGTAAGAACTCATTAAGTAAGTAATTC 1080
OY 1081 TGCTGTGATATCAATATATCTCAGATGTATTCATCTAATCAAGATTTTTCATATCTCA 1140
Db 1081 TGCTGTGATATCAATATATCTCAGATGTATTCATCTAATCAAGATTTTTCATATCTCA 1140
OY 1141 ACTAGATATCTTTAGACATGCTTAAATATCAAAAGCAGTTGATTTGGAAGTACATT 1200
Db 1141 ACTAGATATCTTTAGACATGCTTAAATATCAAAAGCAGTTGATTTGGAAGTACATT 1200

Db 1141 ACTAGATATCTTTAGACATGCTTAAATATCAAAAGCAGTTGATTTGGAAGTACATT 1200
OY 1201 GTGATAGATGTGCAAGGAGACATATTTGATGATGATGATGATGATGATGATGAT 1260
Db 1201 GTGATAGATGTGCAAGGAGACATATTTGATGATGATGATGATGATGATGATGAT 1260
OY 1261 AAAATTTATTTGCTG 1275
Db 1261 AAAATTTATTTGCTG 1275
RESULT 3
US-08-609-443B-14/C
Sequence 14, Application US/08609443B
Patent No. 5840693
GENERAL INFORMATION:
APPLICANT: ERIKSSON, Ulf
APPLICANT: OLOFSSON, Birgitta
APPLICANT: ALITALO, Kari
APPLICANT: PAISOLA, Kari
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W., Suite 700
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/609,443B
FILING DATE: 01-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/397,651
FILING DATE: 01-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,427
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569,063
FILING DATE: 06-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: EVANS, Joseph D
REGISTRATION NUMBER: 26,269
REFERENCE/DOCKET NUMBER: 1064/41979C4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 624 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
TISSUE TYPE: human
US-08-609-443B-14
Query Match 1.4%, Score 18; DB 2; Length 624;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 585 GGCAGCGCGCGGCGAGG 568

RESULT 4

US-08-569-063C-14/C
Sequence 14, Application US/08569063C
Patent No. 5928939

GENERAL INFORMATION:
APPLICANT: ERIKSSON, Ulf
APPLICANT: OLOFSSON, Birgitta
APPLICANT: ALITALO, Kari

APPLICANT: PAJUSOLA, Katri
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
TITLE OF INVENTION: DNA CODING THEREFOR
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W., Suite 700
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569/063C
FILING DATE: 06-DEC-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,427
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/397,651
FILING DATE: 01-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: EVANS, Joseph D
REGISTRATION NUMBER: 26,269
REFERENCE/DOCKET NUMBER: 1064/41979CP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 624 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:

TISSUE TYPE: human
US-08-569-063C-14

Query Match

Best Local Similarity 1.4%; Score 18; DB 2; Length 624;
Matches 18; Conservativity 100.0%; Pred. No. 12;
0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGCGCGGCGGCGAGG 77
DB 585 GGCAGCGCGCGGCGGCGAGG 568

RESULT 5

5459251-1/c

PATENT NO. 5459251
APPLICANT: Tsujimoto, Yoshida, Croce, Carlo A.
TITLE OF INVENTION: DNA MOLECULES HAVING HUMAN BCL-2 GENE
SEQUENCES

NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/228,704

FILING DATE: 18-APR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 994,941
FILING DATE: 23-DEC-1992
APPLICATION NUMBER: 663,010
FILING DATE: 19-MAR-1991
APPLICATION NUMBER: 883,687
FILING DATE: 09-JUL-1986

SEQ ID NO: 1
LENGTH: 4825
5459251-1

Query Match

Best Local Similarity 1.4%; Score 18; DB 6; Length 4825;
Matches 18; Conservativity 100.0%; Pred. No. 13;
0; Mismatches 0; Indels 0; Gaps 0;

QY 59 TGGCAGCGCGCGGCGGCGAG 76
DB 694 TGGCAGCGCGCGGCGGCGAG 677

RESULT 6

US-08-465-485A-19/C
Sequence 19, Application US/08465485A
Patent No. 5831066

GENERAL INFORMATION:
APPLICANT: Reed, John
TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, WATERS & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Hwy., Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,485A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/124,256
FILING DATE: 20-SEP-1993

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
US-08-465-485A-19

Query Match

Best Local Similarity 1.4%; Score 18; DB 2; Length 5086;
Matches 18; Conservativity 100.0%; Pred. No. 13;
0; Mismatches 0; Indels 0; Gaps 0;

QY 59 TGGCAGCGCGCGGCGGCGAG 76
DB 694 TGGCAGCGCGCGGCGGCGAG 677

RESULT 6

US-08-465-485A-19/C

PATENT NO. 5831066
APPLICANT: Reed, John
TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, WATERS & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Hwy., Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,485A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/124,256
FILING DATE: 20-SEP-1993
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-465-485A-19

Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 TGGCAGCGCGCGCGCAG 76
|||||
Db 694 TGGCAGCGCGCGCGCAG 677

RESULT 7

US-08-365-486A-14/C
Sequence 14, Application US/08365486A
Patent No. 5834306
GENERAL INFORMATION:
APPLICANT: Webster, Keith A.
TITLE OF INVENTION: Tissue Specific Hypoxia Regulated
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/365,486A
FILING DATE: 23-DEC-1994
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 8255-0018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: human bcl-2 cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1459..2178
US-08-365-486A-14

Query Match 1.4%; Score 18; DB 2; Length 5086;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 TGGCAGCGCGCGCGCAG 76
|||||
Db 694 TGGCAGCGCGCGCGCAG 677

RESULT 8
US-09-080-285-19/C
Sequence 19, Application US/09080285
Patent No. 6040181
GENERAL INFORMATION:
APPLICANT: Reed, John

TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Hwy., Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,285
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/465,485
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/124,256
FILING DATE: 20-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/840,716
FILING DATE: 21-FEB-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/288,692
FILING DATE: 22-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: Fortney, Andrew D.
REGISTRATION NUMBER: 34,600
REFERENCE/DOCKET NUMBER: 3335-070-55 CONT.

TELECOMMUNICATION INFORMATION:
TELEPHONE: (408) 436-2070
TELEFAX: (408) 436-2075
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-080-285-19

Query Match 1.4%; Score 18; DB 3; Length 5086;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 TGGCAGCGCGCGCGCAG 76
|||||
Db 694 TGGCAGCGCGCGCGCAG 677

RESULT 9

US-08-880-342-14/C
Sequence 14, Application US/08880342
Patent No. 6218179
GENERAL INFORMATION:
APPLICANT: Webster, Keith A.
APPLICANT: Bishopric, Nanette H.
APPLICANT: Murphy, Brian
APPLICANT: Laderoute, Keith R.
APPLICANT: Green, Christopher J.
TITLE OF INVENTION: Tissue Specific Hypoxia Regulated
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto

STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/880,342
FILING DATE: 23-JUN-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/IB95/00996
FILING DATE: 13-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/365,486
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 8255-0018.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: human bcl-2 cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1459..2178
US-08-880-342-14

Query Match
Best Local Similarity: 1.4%; Score 18; DB 4; Length 5086;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 TGGCAGCGCGCGCGCAG 76
|||||
DB 694 TGGCAGCGCGCGCGCAG 677

RESULT 10
PCT-US93-05651-4/c
Sequence 4, Application PC/TUS9305651
GENERAL INFORMATION:
TITLE OF INVENTION: A Gene which Prevents Programmed Cell Death
NUMBER OF SEQUENCES: 5
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05651
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1459..2178
OTHER INFORMATION: /product="bcl-2"
PCT-US93-05651-4

Query Match
Best Local Similarity: 1.4%; Score 18; DB 5; Length 5086;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 TGGCAGCGCGCGCGCAG 76
|||||
DB 694 TGGCAGCGCGCGCGCAG 677

RESULT 11
PCT-US93-06251-2/c
Sequence 2, Application PC/TUS9306251
GENERAL INFORMATION:
APPLICANT: Wickstrom, Eric and Rife, Jason P.
TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06251
FILING DATE: 19930630
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8586
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-2

Query Match
Best Local Similarity: 1.4%; Score 18; DB 5; Length 5086;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 TGGCAGCGCGCGCGCAG 76
|||||
DB 694 TGGCAGCGCGCGCGCAG 677

RESULT 12
5506344-1/c
Patent No. 5506344
APPLICANT: TSUJIMOTO, YOSHIHIDE;CROCE, CARLO A.
TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR BCL-2GENE PRODUCT
NUMBER OF SEQUENCES: 5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,193
FILING DATE: 05-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 228,704

FILING DATE: 18-APR-1994
APPLICATION NUMBER: 994,941
FILING DATE: 23-DEC-1992
APPLICATION NUMBER: 663,010
FILING DATE: 19-MAR-1991
APPLICATION NUMBER: 883,687
FILING DATE: 09-JUL-1986
SEQ ID NO: 1
LENGTH: 5104
5506344-1

Query Match 1.4%; Score 18; DB 6; Length 5104;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 TGGCAGCGCGCGCGCAG 76
|||||
Db 694 TGGCAGCGCGCGCGCAG 677

RESULT 13
US-08-180-524-5/c
Sequence 5, Application US/08180524
Patent No. 5849537

GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Huige, Nick
APPLICANT: Kot, Edward
APPLICANT: Chicoye, Etzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
TITLE OF INVENTION: IN YEAST
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Thad Kryshak, Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
FILING DATE: 28-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 77 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Oligonucleotide
HYPOTHETICAL: YES
US-08-180-524-5

Query Match 1.3%; Score 17; DB 2; Length 77;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGCGCGCGC 74
|||||
Db 54 GTGGCAGCGCGCGCGC 38

RESULT 14
US-08-975-166-5/c
Sequence 5, Application US/08975166
Patent No. 5928877

GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Huige, Nick
APPLICANT: Kot, Edward
APPLICANT: Chicoye, Etzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
TITLE OF INVENTION: IN YEAST
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Thad Kryshak, Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,166
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
FILING DATE: 28-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 77 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Oligonucleotide

HYPOTHETICAL: YES
US-08-975-166-5

Query Match 1.3%; Score 17; DB 2; Length 77;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGCGCGCGC 74
|||||
DB 54 GTGGCAGCGCGCGCGC 38

RESULT 15
US-08-180-524-3/C

Sequence 3, Application US/08180524
Patent No. 5849537

GENERAL INFORMATION:

APPLICANT: Tripp, Matthew

APPLICANT: Lusk, Lance

APPLICANT: Rhodes, Thomas

APPLICANT: Huige, Nick

APPLICANT: Kot, Edward

APPLICANT: Chicoye, Elzer

APPLICANT: Barney, Michael C.

APPLICANT: Bower, Patricia A.

APPLICANT: Cronan, Charles L.

TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS

TITLE OF INVENTION: IN YEAST

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Thad Kryshak, Quarles & Brady

STREET: 411 East Wisconsin Avenue

CITY: Milwaukee

STATE: Wisconsin

COUNTRY: USA

ZIP: 53202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MSWORD Version 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/180,524

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/917,216

FILING DATE:

APPLICATION NUMBER: US 07/486,333

FILING DATE: 28-FEB-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/409,217

FILING DATE: 19-SEP-1989

ATTORNEY/AGENT INFORMATION:

NAME: Kryshak, Thad

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 66-005-9234-1

TELEPHONE: (414) 277-5707

TELEFAX: (414) 277-5591

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 146 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Other Nucleic Acid

HYPOTHETICAL: YES

ANTI-SENSE: NO

Query Match 1.3%; Score 17; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGCGCGCGC 74
|||||
DB 123 GTGGCAGCGCGCGCGC 107

RESULT 16
US-08-975-166-3/C

Sequence 3, Application US/08975166
Patent No. 5928877

GENERAL INFORMATION:

APPLICANT: Tripp, Matthew

APPLICANT: Lusk, Lance

APPLICANT: Rhodes, Thomas

APPLICANT: Huige, Nick

APPLICANT: Kot, Edward

APPLICANT: Chicoye, Elzer

APPLICANT: Barney, Michael C.

APPLICANT: Bower, Patricia A.

APPLICANT: Cronan, Charles L.

TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS

TITLE OF INVENTION: IN YEAST

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Thad Kryshak, Quarles & Brady

STREET: 411 East Wisconsin Avenue

CITY: Milwaukee

STATE: Wisconsin

COUNTRY: USA

ZIP: 53202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MSWORD Version 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/975,166

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/180,524

FILING DATE:

APPLICATION NUMBER: US/07/917,216

FILING DATE:

APPLICATION NUMBER: US 07/486,333

FILING DATE: 28-FEB-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/409,217

FILING DATE: 19-SEP-1989

ATTORNEY/AGENT INFORMATION:

NAME: Kryshak, Thad

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 66-005-9234-1

TELEPHONE: (414) 277-5707

TELEFAX: (414) 277-5591

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 146 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Other Nucleic Acid

HYPOTHETICAL: YES

ANTI-SENSE: NO

Query Match 1.3%; Score 17; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 35;

APPLICANT: Ward, Judith
TITLE OF INVENTION: 91dB
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/097,072
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/886,638
FILING DATE: 01-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10028-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 810 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-097-072-1

Query Match 1.3%; Score 17; DB 3; Length 810;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 TACATCGTTCATCTG 1007
|||||
DB 203 TACATCGTTCATCTG 187

RESULT 21
US-08-886-638-1/c
Sequence 1, Application US/08886638
Patent No. 5866367
GENERAL INFORMATION:
APPLICANT: Kallender, Howard
TITLE OF INVENTION: NOVEL COMPOUNDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 997 Lenox Drive, Building 3, Suite 210
CITY: Lawrenceville
STATE: NJ
COUNTRY: USA
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/886,638
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: GM10028
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 900 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-886-638-1

Query Match 1.3%; Score 17; DB 2; Length 900;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 TACATCGTTCATCTG 1007
|||||
DB 221 TACATCGTTCATCTG 205

RESULT 22
US-09-097-072-3/c
Sequence 3, Application US/09097072
Patent No. 6140079
GENERAL INFORMATION:
APPLICANT: Kallender, Howard
APPLICANT: Lennox, Anna L.
APPLICANT: Burdham, Martin K. R.
TITLE OF INVENTION: 91dB
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/097,072
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/886,638
FILING DATE: 01-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10028-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 900 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-097-072-3

Query Match 1.3% Score 17; DB 3; Length 900;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 TACATCGTGTTCATCTG 1007
|||||
DB 221 TACATCGTGTTCATCTG 205

RESULT 23

US-09-212-988-1/c
Sequence 1, Application US/09212988
Patent No. 6210673
GENERAL INFORMATION:
APPLICANT: kallender, Howard
TITLE OF INVENTION: NOVEL COMPOUNDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 997 Lenox Drive, Building 3, Suite 210
CITY: Lawrenceville
STATE: NJ
COUNTRY: USA
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/212.988
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/886,638
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: GM10028
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 900 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-212-988-1

Query Match 1.3% Score 17; DB 4; Length 900;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 TACATCGTGTTCATCTG 1007
|||||
DB 221 TACATCGTGTTCATCTG 205

RESULT 24
PCT-US91-00899-13
Sequence 13, Application PC/TUS9100899
GENERAL INFORMATION:
APPLICANT: Lowe, John B.
TITLE OF INVENTION: Method and Products For the Synthesis of
Oligosaccharide Structures on Glycoproteins, Glycolipids,
or as Free Molecules, and For the Isolation of Cloned
Genetic Sequences That Determine These Structures

NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/00899
FILING DATE: 19910214
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavallee Ph.D., Jean-Paul
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-021-55 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-5940
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1068 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
PCT-US91-00899-13

Query Match 1.3% Score 17; DB 5; Length 1068;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGGCGCGGCAG 76
|||||
DB 294 GGCAGCGGCGCGGCAG 310

RESULT 25
PCT-US91-00899-5
Sequence 5, Application PC/TUS9100899
GENERAL INFORMATION:
APPLICANT: Lowe, John B.
TITLE OF INVENTION: Method and Products For the Synthesis of
Oligosaccharide Structures on Glycoproteins, Glycolipids,
or as Free Molecules, and For the Isolation of Cloned
Genetic Sequences That Determine These Structures
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/00899
FILING DATE: 19910214
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavallee Ph.D., Jean-Paul

REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-021-55 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-5940
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1215 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
PCT-US91-00899-5

Query Match 1.3%; Score 17; DB 5; Length 1215;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGCGCGCGCAG 76
|||||
DB 441 GGCAGCGCGCGCGCAG 457

RESULT 26
US-07-914-281-9
; Sequence 9, Application US/07914281
; Patent No. 5324663
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; NUMBER OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
; ADDRESSEE: P.C.,
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/914,281
; FILING DATE: 19920720
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1488 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-07-914-281-9

Query Match 1.3%; Score 17; DB 1; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGCGCGCGCAG 76
|||||
DB 714 GGCAGCGCGCGCGCAG 730

RESULT 27
US-08-393-246-9
; Sequence 9, Application US/08393246
; Patent No. 5595900
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; NUMBER OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
; ADDRESSEE: P.C.,
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,246
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/220,433
; FILING DATE: 30-MAR-1994
; APPLICATION NUMBER: US 07/914,281
; FILING DATE: 20-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1488 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-393-246-9

Query Match 1.3%; Score 17; DB 1; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGCGCGCGCAG 76
|||||
DB 714 GGCAGCGCGCGCGCAG 730

RESULT 28
US-08-525-058A-9
; Sequence 9, Application US/08525058A
; Patent No. 5770420
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
TITLE OF INVENTION: 435
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavallee, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-525-058A-9

Query Match 1.38; Score 17; DB 1; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 60 GGCAGCGCGCGGCAG 76
|||||
Db 714 GGCAGCGCGCGGCAG 730

RESULT 29
US-08-696-731-9
Sequence 9, Application US/08696731
Patent No. 595347
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
TITLE OF INVENTION: 14
CLASSIFICATION: 14
ATTORNEY/AGENT INFORMATION:
NAME: Lavallee, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-525-058A-9

Query Match 1.38; Score 17; DB 1; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 60 GGCAGCGCGCGGCAG 76
|||||
Db 714 GGCAGCGCGCGGCAG 730

APPLICATION NUMBER: US/08/696,731
FILING DATE: 14-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavallee, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-696-731-9

Query Match 1.38; Score 17; DB 2; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 60 GGCAGCGCGCGGCAG 76
|||||
Db 714 GGCAGCGCGCGGCAG 730

RESULT 30
US-09-042-531-9
Sequence 9, Application US/09042531
Patent No. 6268193
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
TITLE OF INVENTION: 14
CLASSIFICATION: 14
ATTORNEY/AGENT INFORMATION:
NAME: Lavallee, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-696-731-9

Query Match 1.38; Score 17; DB 2; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 60 GGCAGCGCGCGGCAG 76
|||||
Db 714 GGCAGCGCGCGGCAG 730

NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-09-042-531-9

Query Match 1.3%; Score 17; DB 4; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCAG 76
|||||
DB 714 GGCAGCGCGCGCGCAG 730

RESULT 31
US-08-365-486A-16/c
Sequence 16, Application US/08365486A
Patent No. 5834306
GENERAL INFORMATION:
APPLICANT: Webster, Keith A.
APPLICANT: Bishopric, Nanette H.
TITLE OF INVENTION: Tissue Specific Hypoxia Regulated
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESS: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/365,486A
FILING DATE: 23-DEC-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 8255-0018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0960
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1846 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: bcl-2 fusion gene; Seto, et al.,
INDIVIDUAL ISOLATE: EMBO J 7:123 (1988)
FEATURE:
NAME/KEY: CDS

LOCATION: 887..1606
US-08-365-486A-16

Query Match 1.3%; Score 17; DB 2; Length 1846;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCAG 76
|||||
DB 339 GGCAGCGCGCGCGCAG 323

RESULT 32
US-08-880-342-16/c
Sequence 16, Application US/08880342
Patent No. 6218179
GENERAL INFORMATION:
APPLICANT: Webster, Keith A.
APPLICANT: Bishopric, Nanette H.
APPLICANT: Murphy, Brian
APPLICANT: Laderoute, Keith R.
APPLICANT: Green, Christopher J.
TITLE OF INVENTION: Tissue Specific Hypoxia Regulated
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESS: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/880,342
FILING DATE: 23-JUN-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/IB95/00996
FILING DATE: 13-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/365,486
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 8255-0018,30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0960
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1846 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: bcl-2 fusion gene; Seto, et al.,
INDIVIDUAL ISOLATE: EMBO J 7:123 (1988)
FEATURE:
NAME/KEY: CDS
LOCATION: 887..1606
US-08-880-342-16

Query Match 1.38; Score 17; DB 4; Length 1846;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGCGCGCGCAG 76
Db 339 GGCAGCGCGCGCGCAG 323

RESULT 33

US-08-483-151-3
Sequence 3, Application US/08483151
Patent No. 5858752
GENERAL INFORMATION:
APPLICANT: Seed, Brian
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,151
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Lech, Karen F.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/278001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2134 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-483-151-3

Query Match 1.38; Score 17; DB 2; Length 2134;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGCGCGCGCAG 76
Db 462 GGCAGCGCGCGCGCAG 478

RESULT 34

PCT-US96-06427-3
Sequence 3, Application PC/TUS9606427
GENERAL INFORMATION:
APPLICANT: The General Hospital Corporation
TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA

COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/06427
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/483,151
FILING DATE: 07-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lech, Karen F.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/278001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2134 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US96-06427-3

Query Match 1.38; Score 17; DB 5; Length 2134;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGCGCGCGCAG 76
Db 462 GGCAGCGCGCGCGCAG 478

RESULT 35

US-08-482-073-9
Sequence 9, Application US/08482073
Patent No. 6307025
GENERAL INFORMATION:
APPLICANT: Hession, Catherine A.
APPLICANT: Lobb, Roy R.
APPLICANT: Goeltz, Susan E.
APPLICANT: Osborn, Laurelee
APPLICANT: Benjamin, Christopher D.
TITLE OF INVENTION: ENDOTHELIAL CELL-LEUKOCYTE ADHESION
TITLE OF INVENTION: MOLECULES (BLAMS) AND MOLECULES INVOLVED IN LEUKOCYTE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,073
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/486,336
FILING DATE: 31-OCT-1990
APPLICATION NUMBER: US 07/608298
FILING DATE: 31-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US 90/02357
FILING DATE: 27-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/452675
FILING DATE: 18-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/359516
FILING DATE: 01-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/354151
FILING DATE: 28-APR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: B124CIP4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2175 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-482-073-9

Query Match 1.3%; Score 17; DB 4; Length 2175;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 60 GGCAGCGCGCGCGCGAG 76
|||||
DB 506 GGCAGCGCGCGCGCGAG 522

RESULT 36
US-08-482-073-10
Sequence 10, Application US/08482073
Patent No. 6307025
GENERAL INFORMATION:
APPLICANT: Hession, Catherine A.
APPLICANT: Lobb, Roy R.
APPLICANT: Goelz, Susan E.
APPLICANT: Osborn, Laurelee
APPLICANT: Benjamin, Christopher D.
APPLICANT: Rosa, Margaret D.
TITLE OF INVENTION: ENDOTHELIAL CELL-LEUKOCYTE ADHESION
TITLE OF INVENTION: MOLECULES (ELAMS) AND MOLECULES INVOLVED IN LEUKOCYTE
NUMBER OF INVENTIONS: ADHESION (MILAS)
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,073
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/486,336
FILING DATE: 31-OCT-1990
APPLICATION NUMBER: US 07/608298
FILING DATE: 31-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US 90/02357
FILING DATE: 27-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/452675
FILING DATE: 18-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/359516
FILING DATE: 01-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/354151
FILING DATE: 28-APR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: B124CIP4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2861 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-482-073-10

Query Match 1.3%; Score 17; DB 4; Length 2861;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 60 GGCAGCGCGCGCGCGAG 76
|||||
DB 989 GGCAGCGCGCGCGCGAG 1005

RESULT 37
US-08-006-676B-2
Sequence 2, Application US/08006676B
Patent No. 5411665
GENERAL INFORMATION:
APPLICANT: Reed, Steven
TITLE OF INVENTION: Diagnosis of Leishmaniasis
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jeffrey B. Oster
STREET: 8339 SE 57th Street
CITY: Mercer Island
STATE: Washington
COUNTRY: USA
ZIP: 98040-4906
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORD for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/006,676B
FILING DATE: 15-JAN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oster, Jeffrey B.
REGISTRATION NUMBER: 32,585
REFERENCE/DOCKET NUMBER: REED-4
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 232 7845
TELEFAX: (206) 236 0205
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE: Leishmania chagasi
US-08-006-676B-2

Query Match 1.3%; Score 17; DB 1; Length 3319;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 CCGAGAGCGCGCGCG 121
|||||
DB 3288 CCGAGAGCGCGCGCG 3304

RESULT 38
US-08-282-845-1
Sequence 1, Application US/08282845
Patent No. 5719263
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: A 230kd Antigen Present in Leishmania
TITLE OF INVENTION: Species
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Macintosh Operating System 7.1
SOFTWARE: Microsoft Word for Macintosh 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/282,845
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/006,676
FILING DATE: JANUARY 15, 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 5004-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: rK39
FEATURE:
NAME/KEY: CDS
LOCATION: 455..3319

US-08-282-845-1

Query Match 1.3%; Score 17; DB 1; Length 3319;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 CCGAGAGCGCGCGCG 121
|||||
DB 3288 CCGAGAGCGCGCGCG 3304

RESULT 39
US-08-428-414A-4
Sequence 4, Application US/08428414A
Patent No. 5912166
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TITLE OF INVENTION: LEISHMANIASIS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,414A
FILING DATE: 21-APR-1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Kadlec, Ann T.
REGISTRATION NUMBER: 39,244
REFERENCE/DOCKET NUMBER: 210121.407
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANDBERRY
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-428-414A-4

Query Match 1.3%; Score 17; DB 2; Length 3319;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 CCGAGAGCGCGCGCG 121
|||||
DB 3288 CCGAGAGCGCGCGCG 3304

RESULT 40
PCT-US94-00324-2
Sequence 2, Application PC/TUS9400324
GENERAL INFORMATION:
APPLICANT: Reed, Steven
TITLE OF INVENTION: Diagnosis of Leishmaniasis
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation

STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple System 7.1
SOFTWARE: Microsoft Word, version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00324
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/006,676
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 5004-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Leishmania chagasi
PCT-US94-00324-2

Query Match 1.3%; Score 17; DB 5; Length 3319;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 CCGAGGAGCGCGCGCG 121
DB 3288 CCGAGGAGCGCGCGCG 3304

RESULT 41
US-08-972-719-1/c
Sequence 1, Application US/08972719
Patent No. 6060588
GENERAL INFORMATION:
APPLICANT: LI, SHENG FENG
APPLICANT: PHILLIPS, DAVID
TITLE OF INVENTION: IDENTIFICATION OF Bap-1, A PROTEIN THAT
BINDS TO INTEGRIN AND IS INVOLVED IN INTEGRIN-MEDIATED
TITLE OF INVENTION: SIGNAL TRANSDUCTION
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVENUE, NW
CITY: WASHINGTON
STATE: CA
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,719
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/753,038
FILING DATE: 19-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, ROBERT A.
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 22803-20011.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFOERSM
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3467 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 226..1233
US-08-972-719-1

Query Match 1.3%; Score 17; DB 3; Length 3467;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1233 ATGTATATGTACCATA 1249
DB 3379 ATGTATATGTACCATA 3363

RESULT 42
US-08-753-038-1/c
Sequence 1, Application US/08753038
Patent No. 6194557
GENERAL INFORMATION:
APPLICANT: LI, SHENG FENG
APPLICANT: PHILLIPS, DAVID
TITLE OF INVENTION: IDENTIFICATION OF Bap-1, A PROTEIN THAT
BINDS TO INTEGRIN AND IS INVOLVED IN INTEGRIN-MEDIATED
TITLE OF INVENTION: SIGNAL TRANSDUCTION
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVENUE, NW
CITY: WASHINGTON
STATE: CA
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/753,038
FILING DATE: 19-NOV-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, ROBERT A.
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 22803-20011.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFOERSM
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3467 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 226..1233
US-08-753-038-1

Query Match 1.3%; Score 17; DB 4; Length 3647;
Best local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1233 ATGTATATGTTACCATA 1249
|||||
DB 3379 ATGTATATGTTACCATA 3363

RESULT 43
US-07-914-281-7
Sequence 7, Application US/07914281
Patent No. 5324663

GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/914,281
FILING DATE: 19920720
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3647 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
US-07-914-281-7

Query Match 1.3%; Score 17; DB 1; Length 3647;
Best local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCGCAG 76
|||||
DB 2382 GGCAGCGCGCGCGCGCAG 2398

RESULT 44

US-08-393-246-7
Sequence 7, Application US/08393246
Patent No. 5595900

GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,246
FILING DATE:
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/220,433
FILING DATE: 30-MAR-1994
APPLICATION NUMBER: US 07/914,281
FILING DATE: 20-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3647 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
US-08-393-246-7

Query Match 1.3%; Score 17; DB 1; Length 3647;
Best local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCGCAG 76
|||||
DB 2382 GGCAGCGCGCGCGCGCAG 2398

RESULT 45
US-08-525-058A-7
Sequence 7, Application US/08525058A
Patent No. 5770420

GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/525,058A
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Lavalleye, Jean-Paul M. P.
 REGISTRATION NUMBER: 31,451
 REFERENCE/DOCKET NUMBER: 2363-060-55
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)521-4500
 TELEFAX: (703)486-2347
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3647 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ANTI-SENSE: NO
 US-08-525-058A-7

Query Match 1.38; Score 17; DB 1; Length 3647;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 60 GGCAGCGGGCGGCGAG 76
 ||||||||||||
 DB 2382 GGCAGCGGGCGGCGAG 2398

Search completed: March 29, 2002, 20:53:54
 Job time: 3524 sec